Exploitation quantitative d’une expérience de physique

Kenneth Maussang

Université de Montpellier

Régression linéaire

On considère un ensemble de N mesures d'une grandeur physique y en fonction d'un paramètre de contrôle x. On note $\{x_i\}$ et $\{y_i\}$ les mesures associées.

Dans le cadre d'une régression linéaire, on suppose une relation linéaire entre les variables x et y de la forme

$$y = \alpha + \beta x.$$

On cherche à déterminer ces valeurs de α et β à partir de valeurs tests a et b (dits ***paramètres d'ajustement***).

On définit **le résidu** ε_i comme l'écart entre le modèle et les mesures expérimentales

$$\varepsilon_i = y_i - a - bx_i.$$

Régression linéaire

La méthode des moindres carrés consiste à choisir pour valeurs de α et β les paramètres a et b minimisant la somme du carré des résidus (***méthode du moindre carré***)

$$\sum_i \varepsilon_i^2 = \sum_i (y_i - a - bx_i)^2\,.$$

On note $\langle X\rangle$, la valeur moyenne sur les N mesures de X

$$\langle X\rangle = \frac{1}{N}\sum_i X_i\,.$$

On définit alors le coefficient de corrélation entre les variables x et y selon

$$r_{xy} = \frac{\langle xy\rangle - \langle x\rangle\langle y\rangle}{\sqrt{(\langle x^2\rangle - \langle x\rangle^2)(\langle y^2\rangle - \langle y\rangle^2)}} = \frac{\langle xy\rangle - \langle x\rangle\langle y\rangle}{\sigma_x\sigma_y}\,,$$

où σ_x et σ_y sont les écart-types des variables x et y.

Régression linéaire

Le **coefficient de détermination** (« coefficient R ») est défini selon

$$R^2 = r_{xy}^2.$$

Le coefficient R prend une valeur entre 0 et 1, et quantifie seulement le degré de corrélation entre les variables x et y.

Quand les variables x et y sont fortement corrélées, $R^2 \to 1$.

Le coefficient de détermination représente le pourcentage de variation de y avec x qui est expliqué par le modèle proposé.

<u>Le coefficient R **ne permet pas** :</u>

- d'indiquer que le modèle proposé est correct ;
- d'exclure un biais expérimental ;
- d'obtenir une estimation des intervalles de confiance des coefficients d'ajustement a et b.

Corrélation ≠ causalité

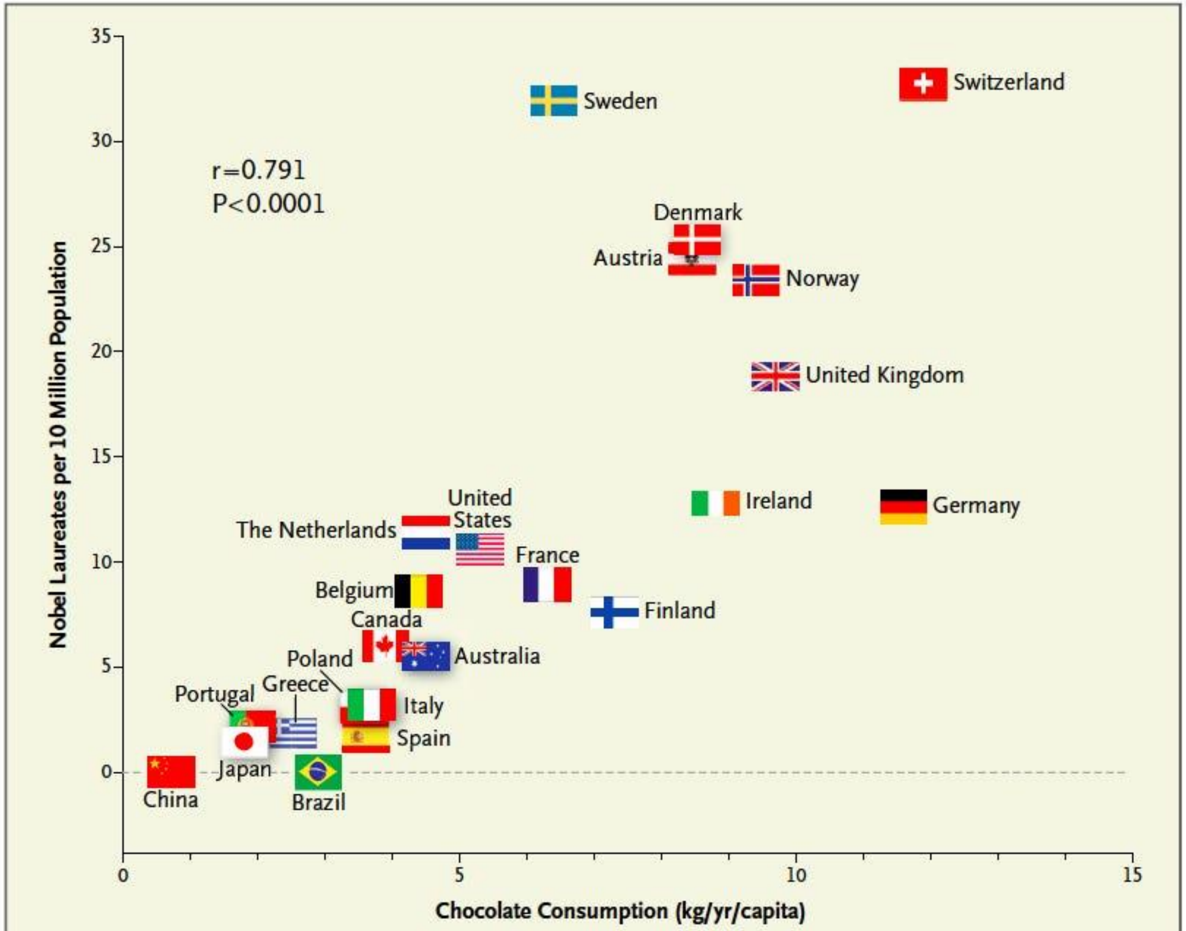
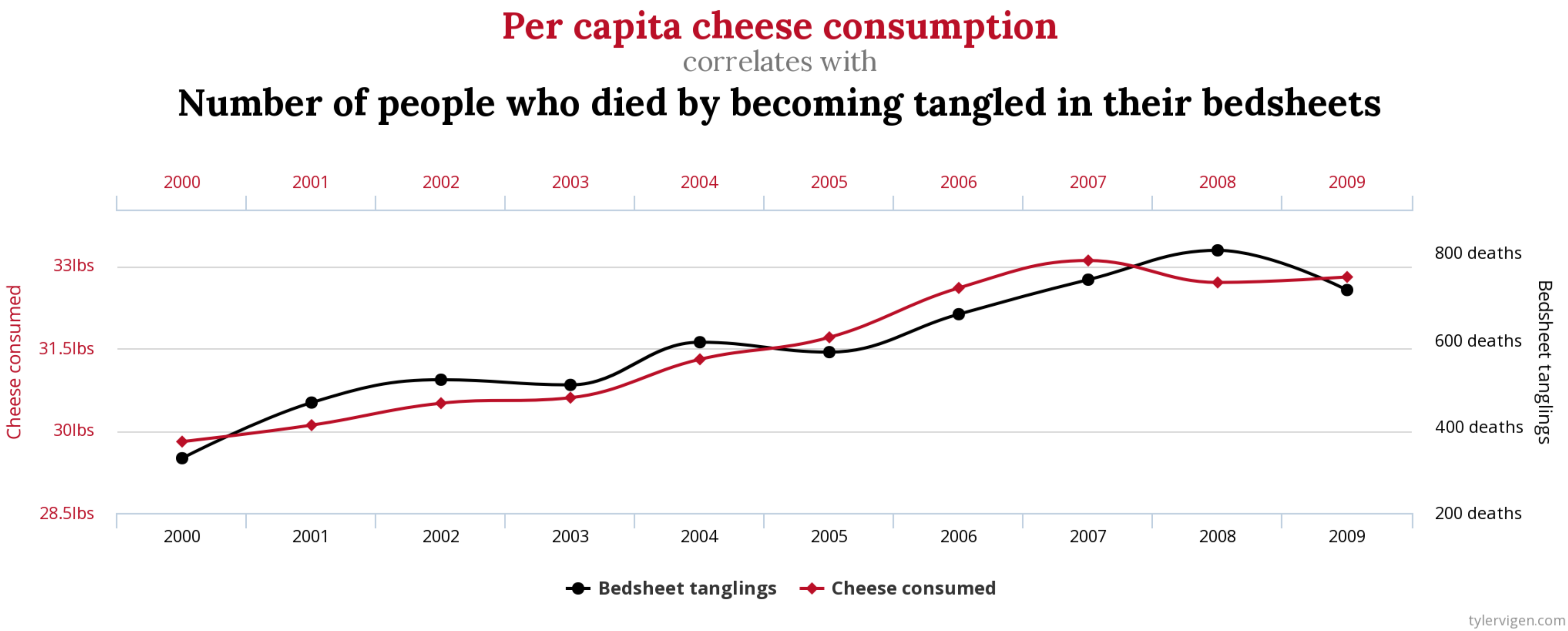


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

https://www.businessinsider.com/chocolate-consumption-vs-nobel-prizes-2014-4?IR=T

Corrélation ≠ causalité



Corrélation : 94,71% !
(R=0,947091)

http://tylervigen.com/spurious-correlations

Comment estimer la qualité d'un ajustement ?

Qualitativement :

en traçant graphiquement les résidus ε_i en fonction du paramètre de contrôle x.

Quantitativement :

en utilisant ***le test du moindre carré***. Cette méthode consiste à vérifier que les résidus sont « distants » de la courbe modèle en raison des incertitudes de mesures. Cela consiste ainsi à normaliser les résidus par rapport à l'incertitude sur la variable y.

Test du moindre carré

Signification de Chi2/(N-p) sous Regressi ?

Ensemble de points expérimentaux $(\{x_i\},\{y_i\})$

Incertitudes pour la variable y_i : σ_i

Fonction de modélisation $y = f(x)$

Nombre de paramètres d'ajustement : p

Nombre de points expérimentaux : N

Définition du paramètre χ^2 :

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - f(x_i))^2}{\sigma_i{}^2} = \sum_{i=1}^{N} \left(\frac{\varepsilon_i}{\sigma_i}\right)^2$$

On définit le χ^2_{red} selon

$$\chi^2_{red} = \frac{\chi^2}{N-p}$$

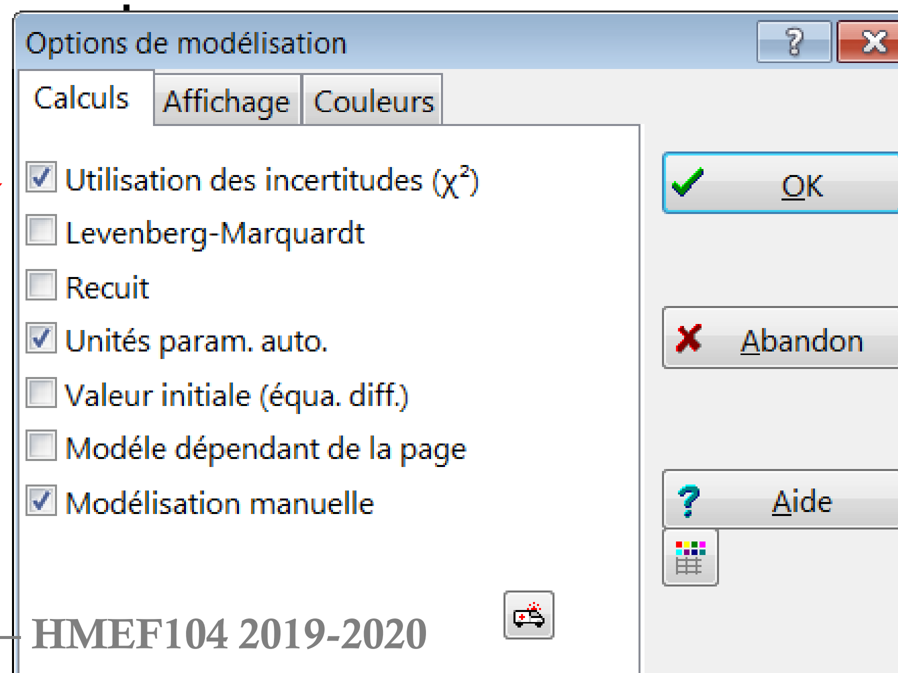
Test du moindre carré

Signification de Chi2/(N-p) sous Regressi ?

Le χ^2 est une mesure de la dispersion quadratique des points expérimentaux à la courbe de modélisation, normalisée à l'incertitude expérimentale de chaque point.

L'algorithme d'ajustement consiste à minimiser le χ^2 afin d'être le plus proche possible des points expérimentaux.

Sous Régressi



Utilisation de Regressi

Signification de Chi2/(N-p) sous Regressi ?

Le χ^2 **réduit** (noté χ^2_{red}) permet de réaliser le ***test des moindres carrés.***

Si les incertitudes sont cohérentes avec la dispersion statistique des points autour de la courbe modèle, on doit avoir

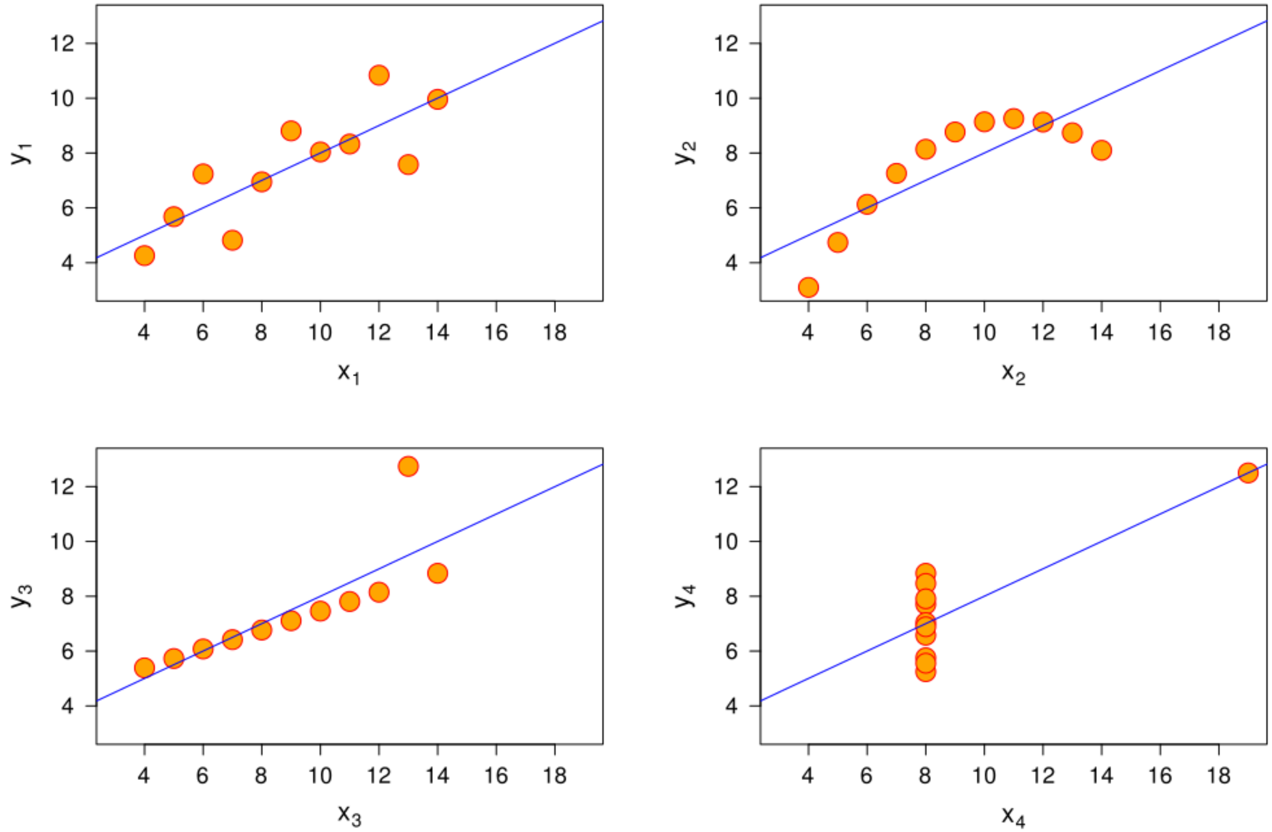
$$\boldsymbol{\chi^2_{red} \sim 1}$$

Si les incertitudes sont sous-estimées, $\chi^2_{red} \gtrsim 1$.

Si les incertitudes sont sur-estimées, $\chi^2_{red} \lesssim 1$.

Exemple : le quartet d'Anscombe

Le quartet d'Anscombe est constitué de quatre ensembles de données qui ont les mêmes propriétés statistiques simples mais qui sont en réalité très différents.



https://fr.wikipedia.org/wiki/Quartet_d%27Anscombe

Exemple : le quartet d’Anscombe

Paramètres | Tableau | Expressions | MathPlayer

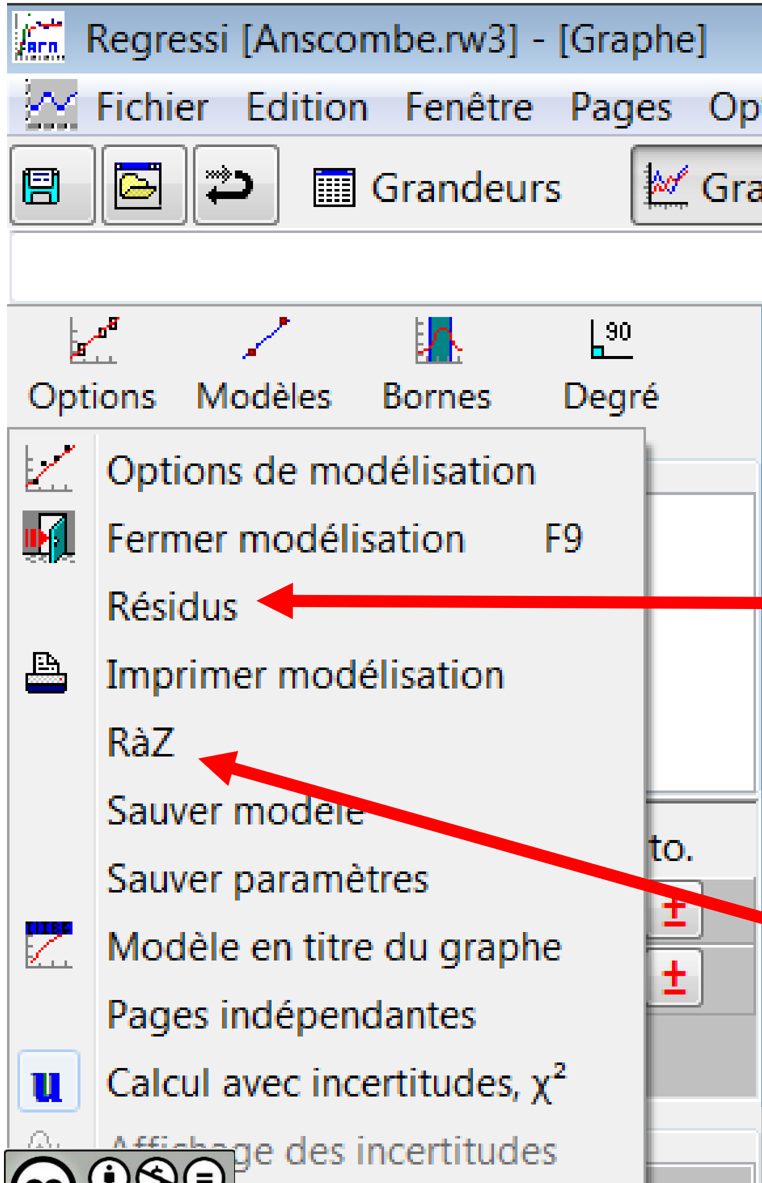
Trier | Ajouter | Sup. colonne | Sup. ligne | Incertitudes | Ajo

i	x	y1	y2	y3	y4	x4
0	4.000	4.260	3.100	5.390	12.50	19.00
1	5.000	5.680	4.740	5.730	6.890	8.000
2	6.000	7.240	6.130	6.080	5.250	8.000
3	7.000	4.820	7.260	6.420	7.910	8.000
4	8.000	6.950	8.140	6.770	5.760	8.000
5	9.000	8.810	8.770	7.110	8.840	8.000
6	10.00	8.040	9.140	7.460	6.580	8.000
7	11.00	8.330	9.260	7.810	8.470	8.000
8	12.00	10.84	9.130	8.150	5.560	8.000
9	13.00	7.580	8.740	12.74	7.710	8.000
10	14.00	9.960	8.100	8.840	7.040	8.000

Les 4 variables y_1, y_2, y_3 et y_4 ont les mêmes propriétés statistiques (valeurs moyennes, écart-type, etc…)

https://fr.wikipedia.org/wiki/Quartet_d%27Anscombe

Régressi : afficher les résidus

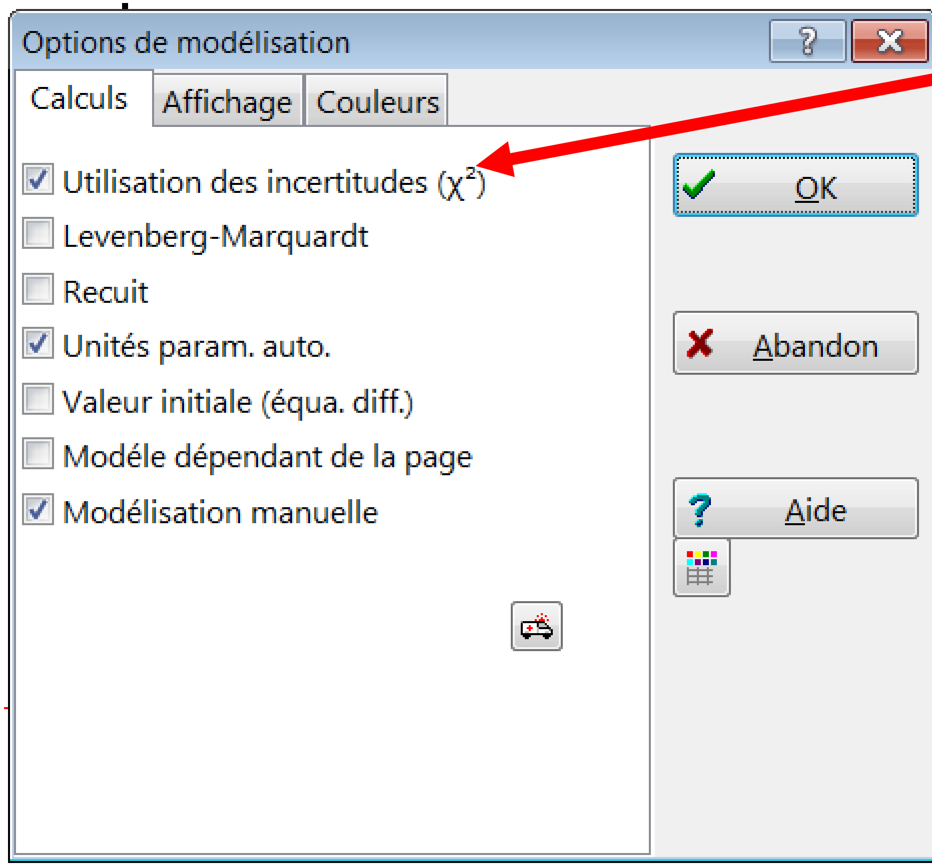


Dans la fenêtre modélisation -> Options

Permet d'afficher les résidus

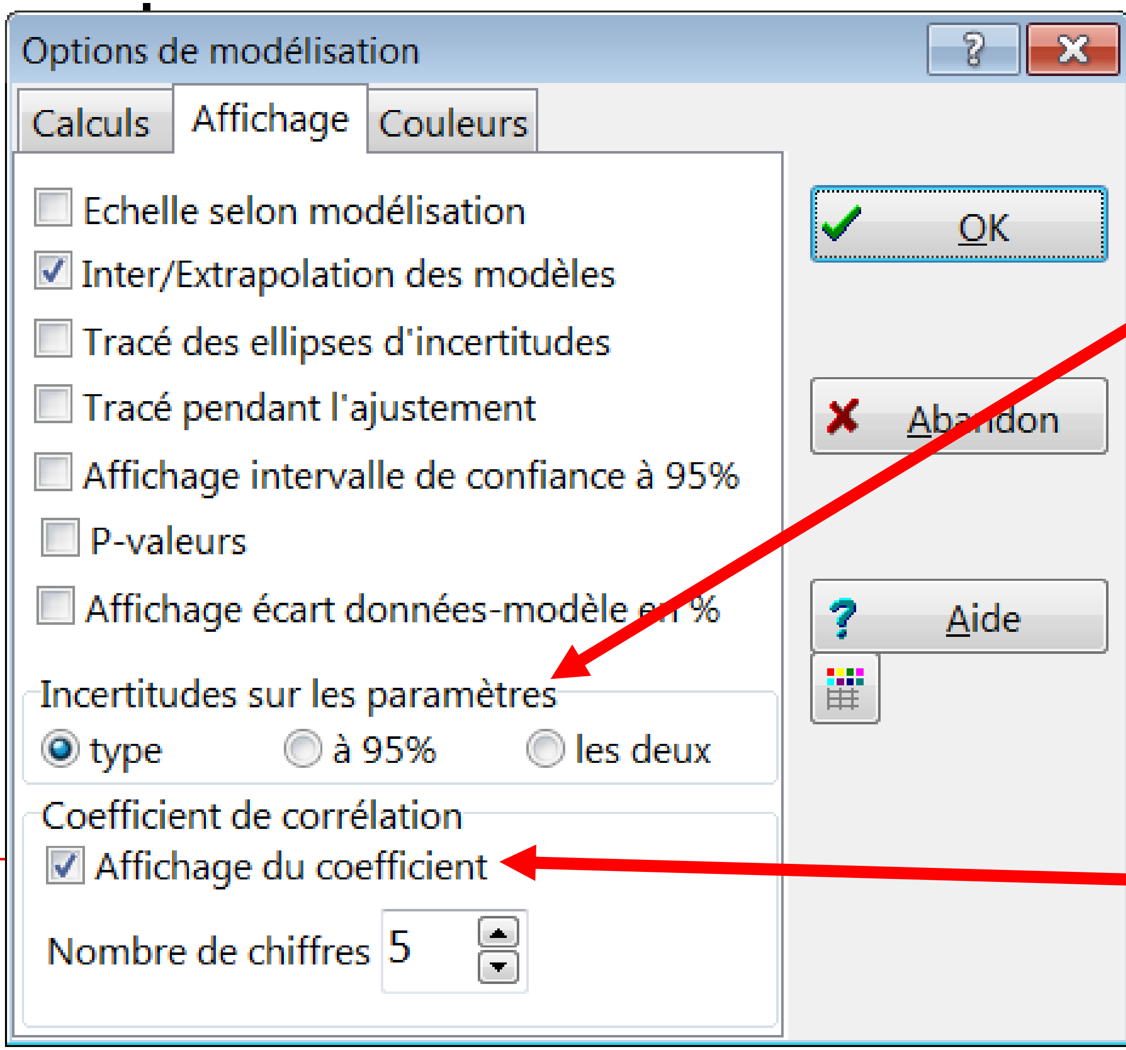
Pour recommencer ou faire un autre ajustement

Régressi : options de modélisation



Permet de tenir compte des incertitudes pour le calcul du moindre carré.

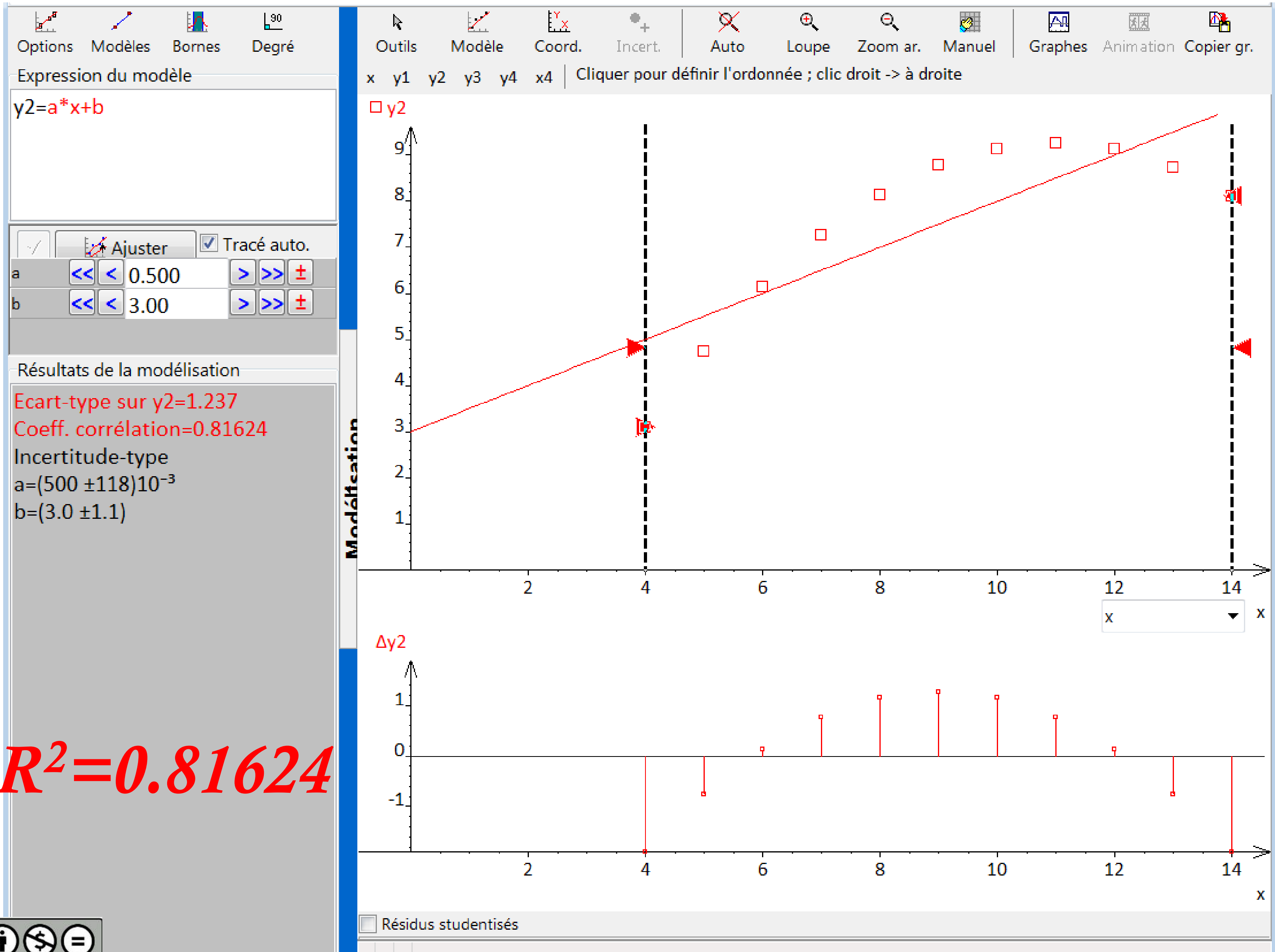
Régressi : options de modélisation



Permet de définir rigoureusement le type d'incertitudes.

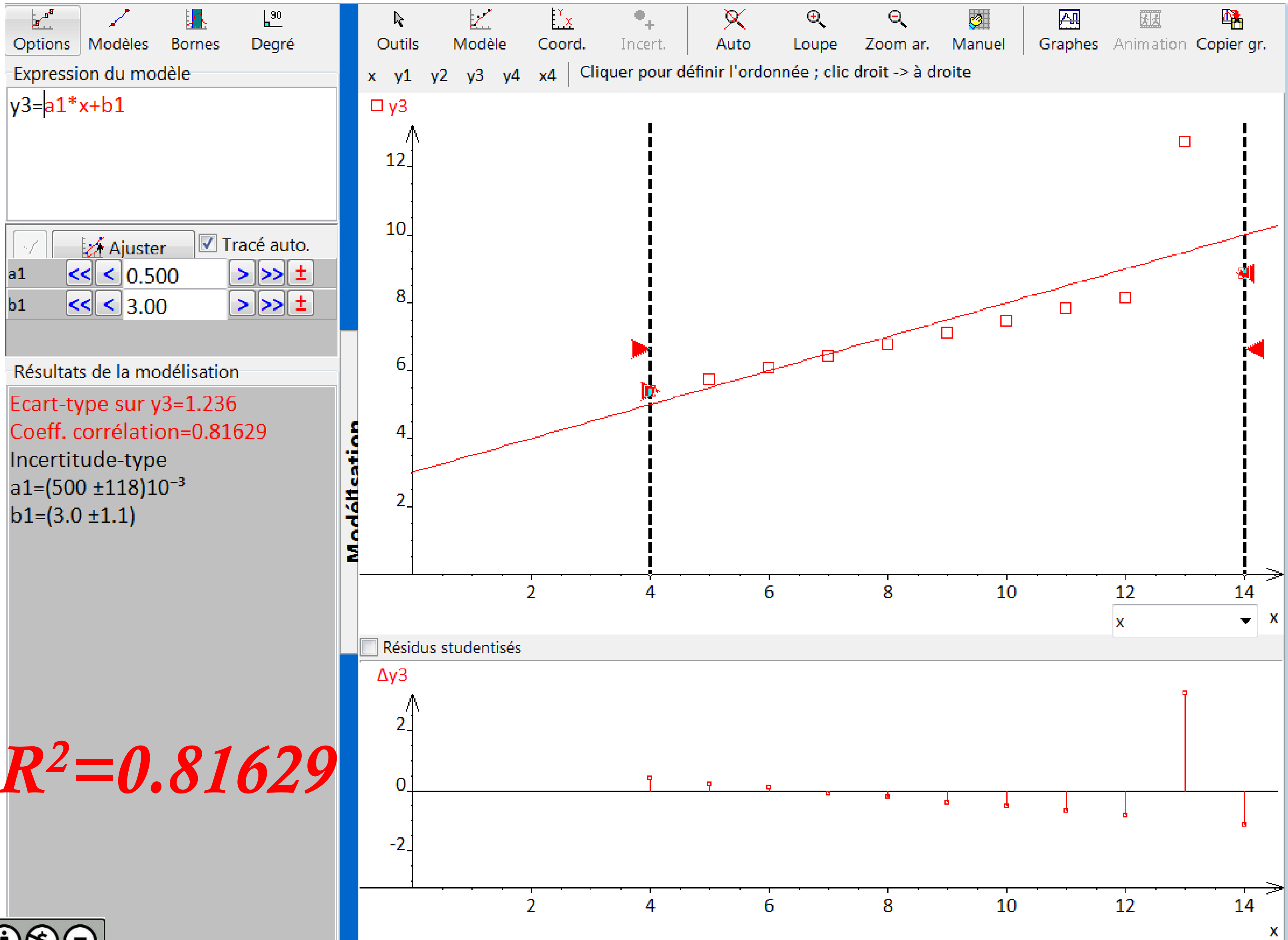
Permet d'afficher la valeur de R^2.

Quartet d'Anscombe : cas de y2



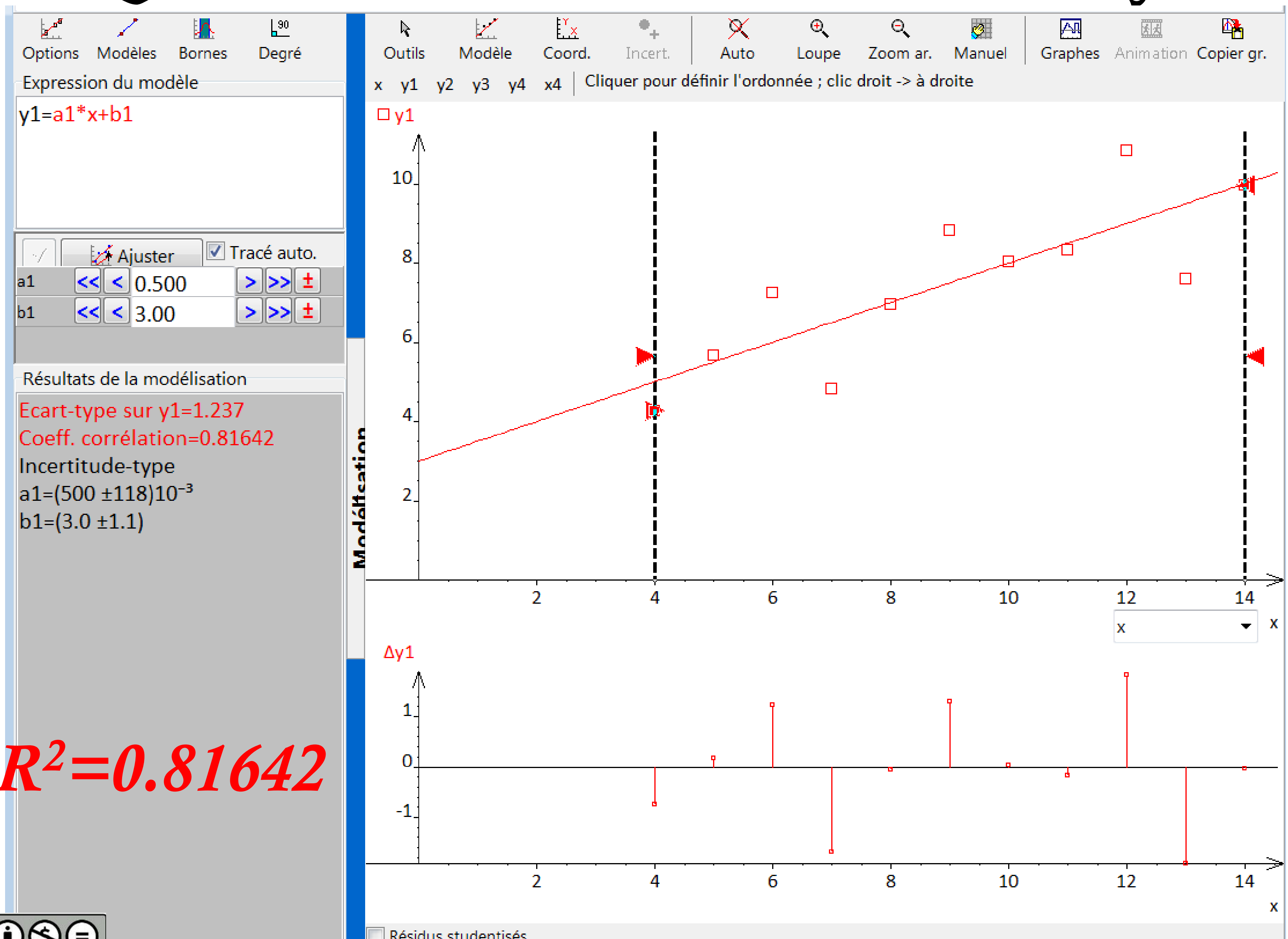
R^2=0.81624

Quartet d'Anscombe : cas de y3



$R^2=0.81629$

Quartet d'Anscombe : cas de y1



R^2=0.81642

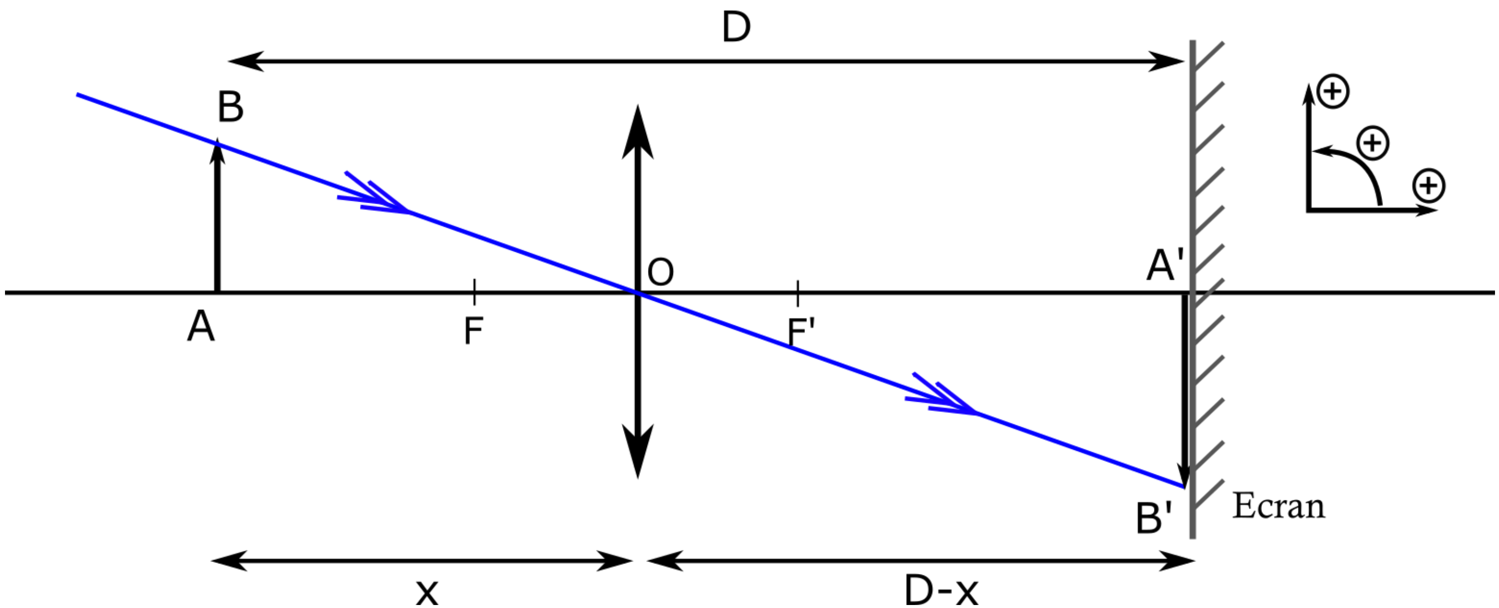
Quartet d'Anscombe : cas de y1

Regressi n'affichera l'information Chi2/(N-p) que si l'on renseigne des valeurs d'incertitudes aux variables.

Il n'est pas possible de faire un test de moindre carré réduit sans incertitudes !

Exemple simple

Mise en application de la méthode de Bessel



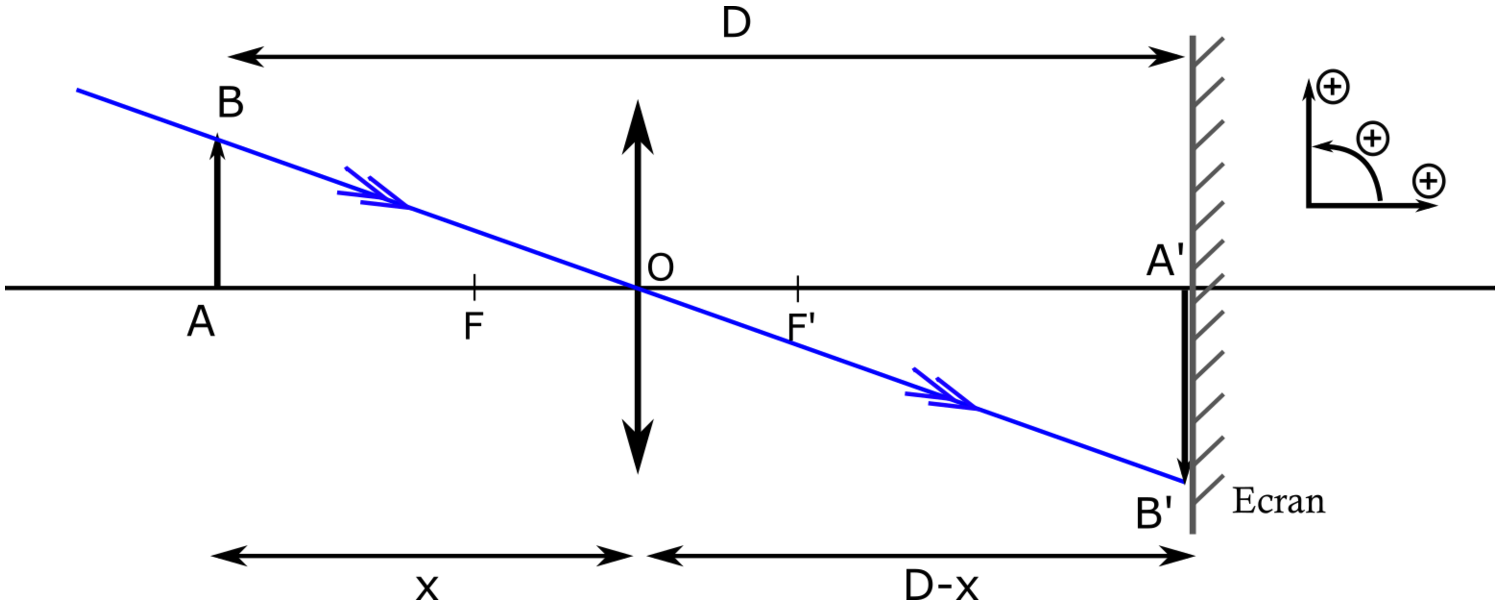
$$x_\pm = \frac{-D \pm \sqrt{D^2 - 4f'D}}{2}$$

$$d = |x_+ - x_-| = \sqrt{D^2 - 4f'D}$$

$$\boxed{f' = \frac{D^2 - d^2}{4D}}$$

Exemple simple

Mise en application de la méthode de Bessel



	Mesure 1	Mesure 2	Mesure 3	Mesure 4
D (mm)	900	1100	1300	1500
x_- (mm)	195	200	187	189
x_+ (mm)	734	955	1142	1360

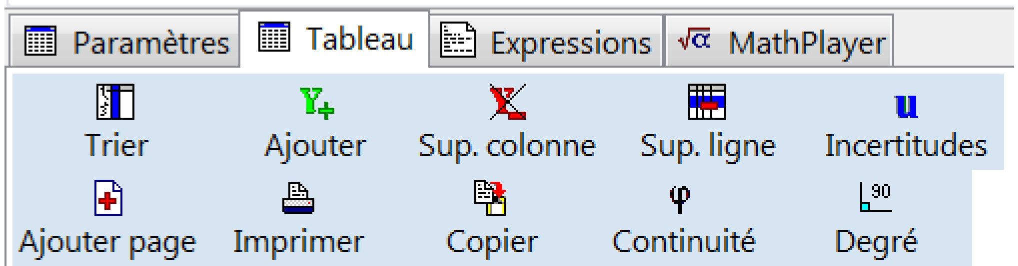
$$\boxed{f' = \frac{D^2 - d^2}{4D}}$$

Incertitude sur D : $\Delta D = 1\text{mm}$

Incertitude sur $x_\pm$: $\Delta x = 4\text{mm}$

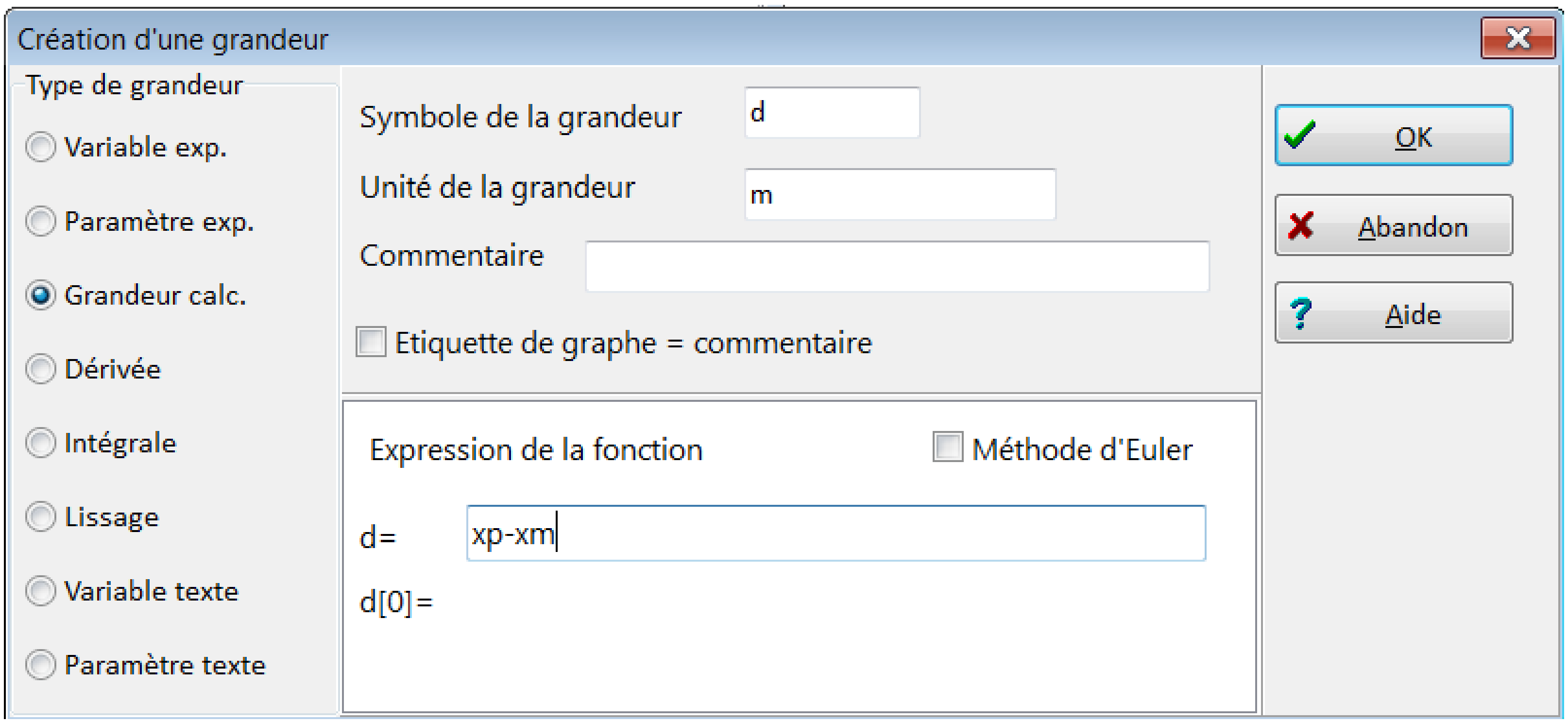
Utilisation de Regressi

$$d = x_+ - x_- \text{ donc } \Delta d = \sqrt{\Delta x^2 + \Delta x^2} = 7\text{mm}$$

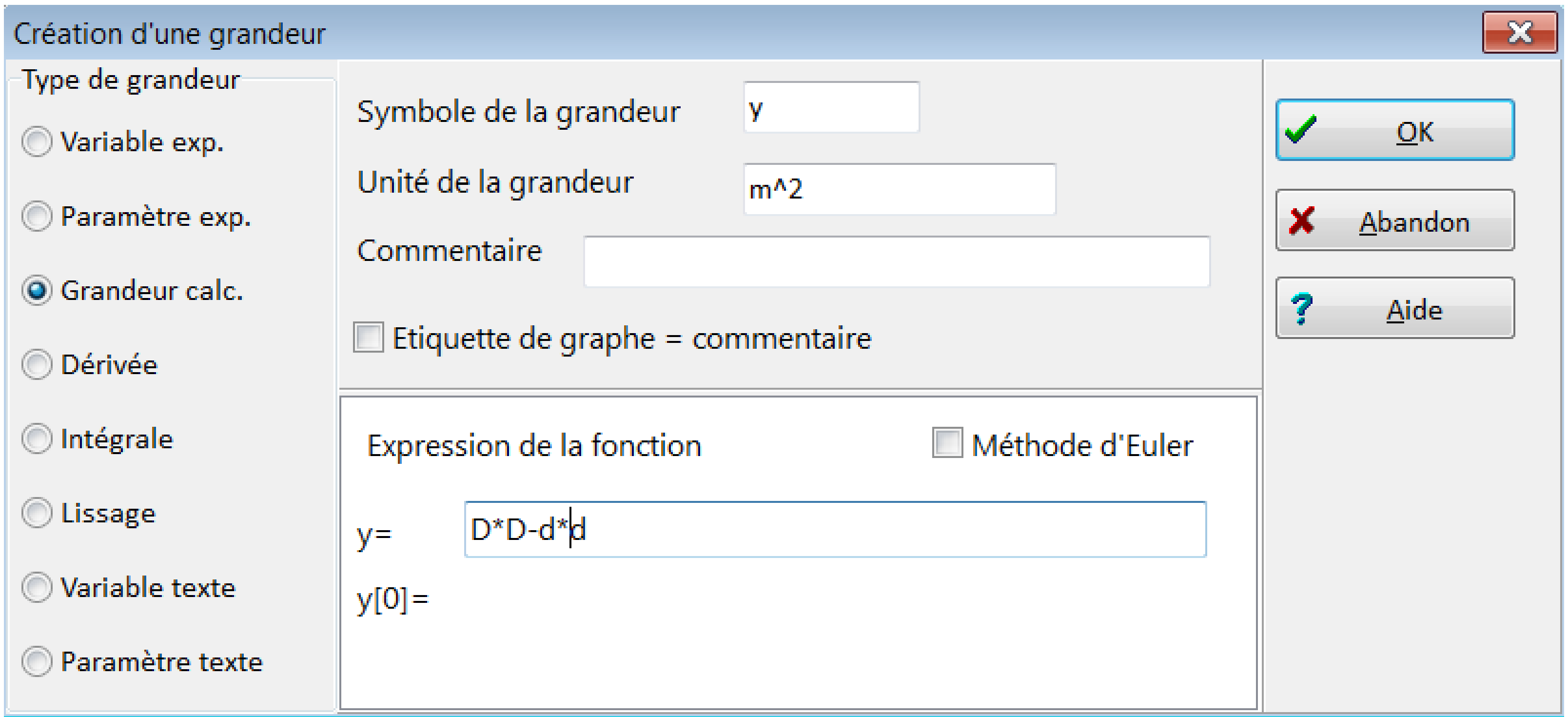


i	D	xp	xm
	m	m	m
0	0.9000	0.7340	0.1950
1	1.100	0.9550	0.2000
2	1.300	1.142	0.1870
3	1.500	1.360	0.1890
4			

Utilisation de Regressi



Utilisation de Regressi



Utilisation de Regressi

Paramètres | Tableau | Expressions | MathPlayer

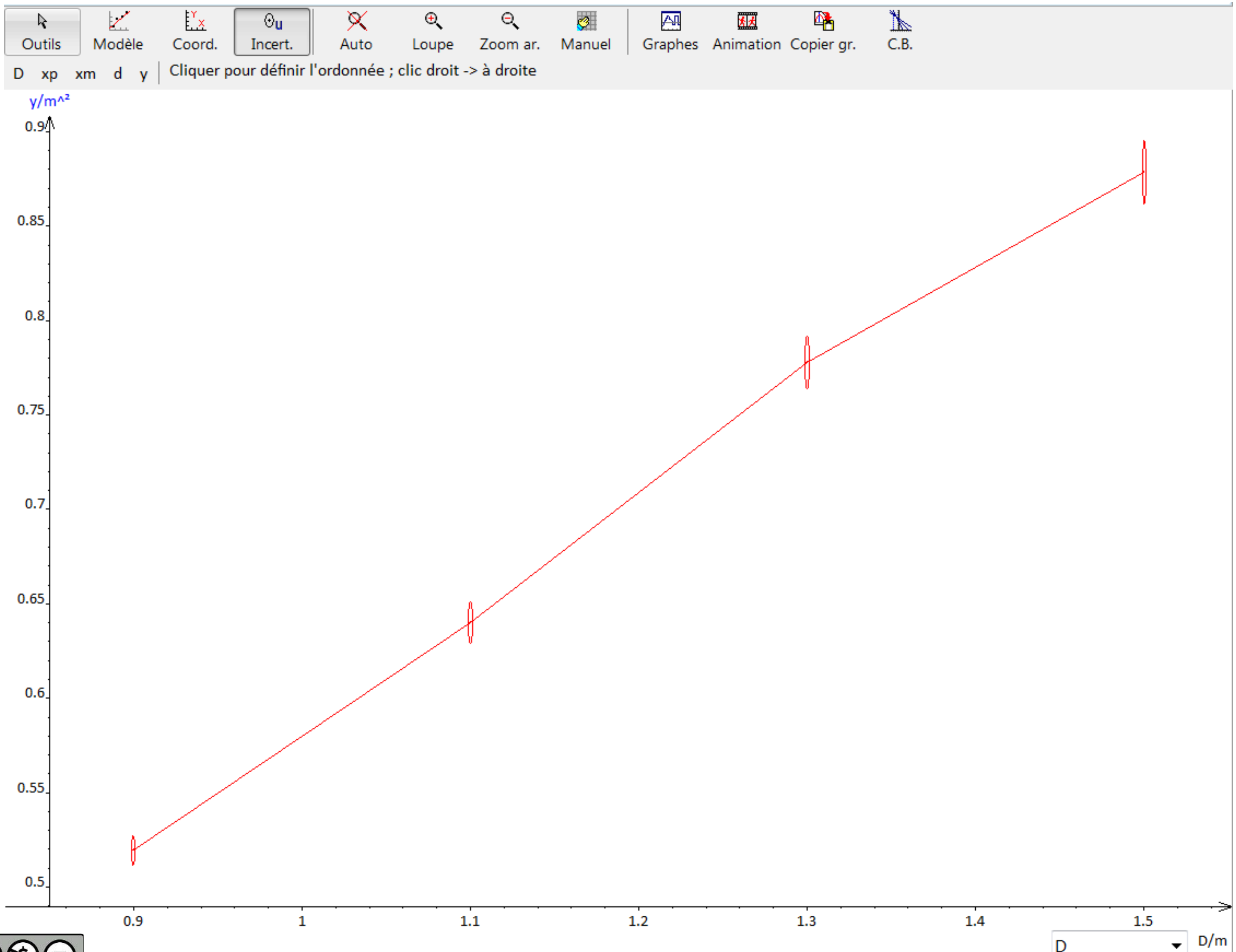
Trier | Ajouter | Sup. colonne | Sup. ligne | Incertitudes | Ajouter page | Imprimer | Copier | Continu

i	D	u(D)	xp	u(xp)	xm	u(xm)	d	u(d)	y	u(y)
	m	m	m	m	m	m	m	m	m^2	m^2
0	0.9000	0.0010	0.7340	0.0040	0.1950	0.0040	0.5390	0.0057	0.5195	0.0064
1	1.100	0.0010	0.9550	0.0040	0.2000	0.0040	0.7550	0.0057	0.6400	0.0088
2	1.300	0.0010	1.142	0.0040	0.1870	0.0040	0.9550	0.0057	0.7780	0.011
3	1.500	0.0010	1.360	0.0040	0.1890	0.0040	1.171	0.0057	0.8788	0.014

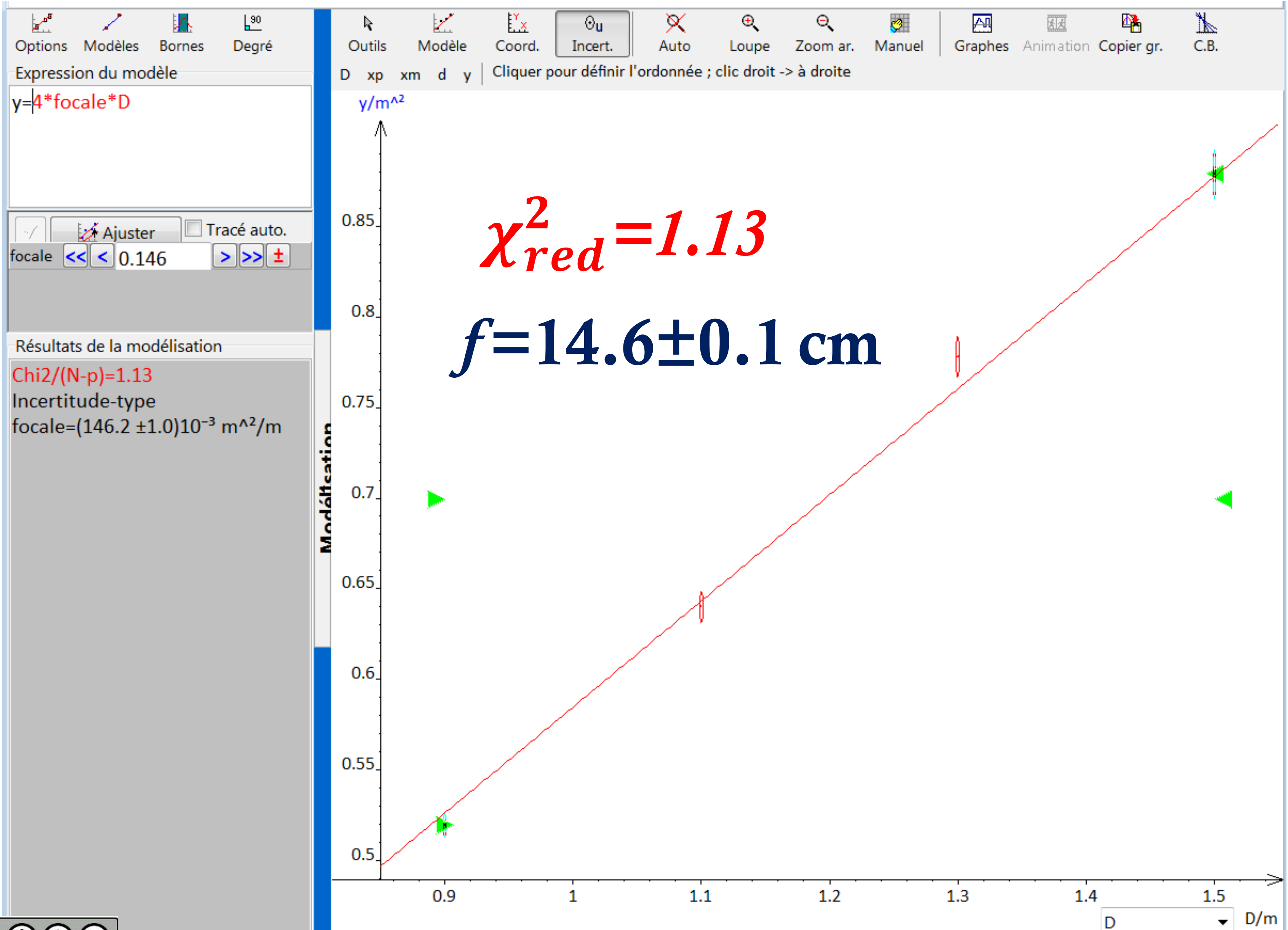
$$\boxed{y = D^2 - d^2 = 4f'D}$$

On trace y en fonction de D.

Utilisation de Regressi



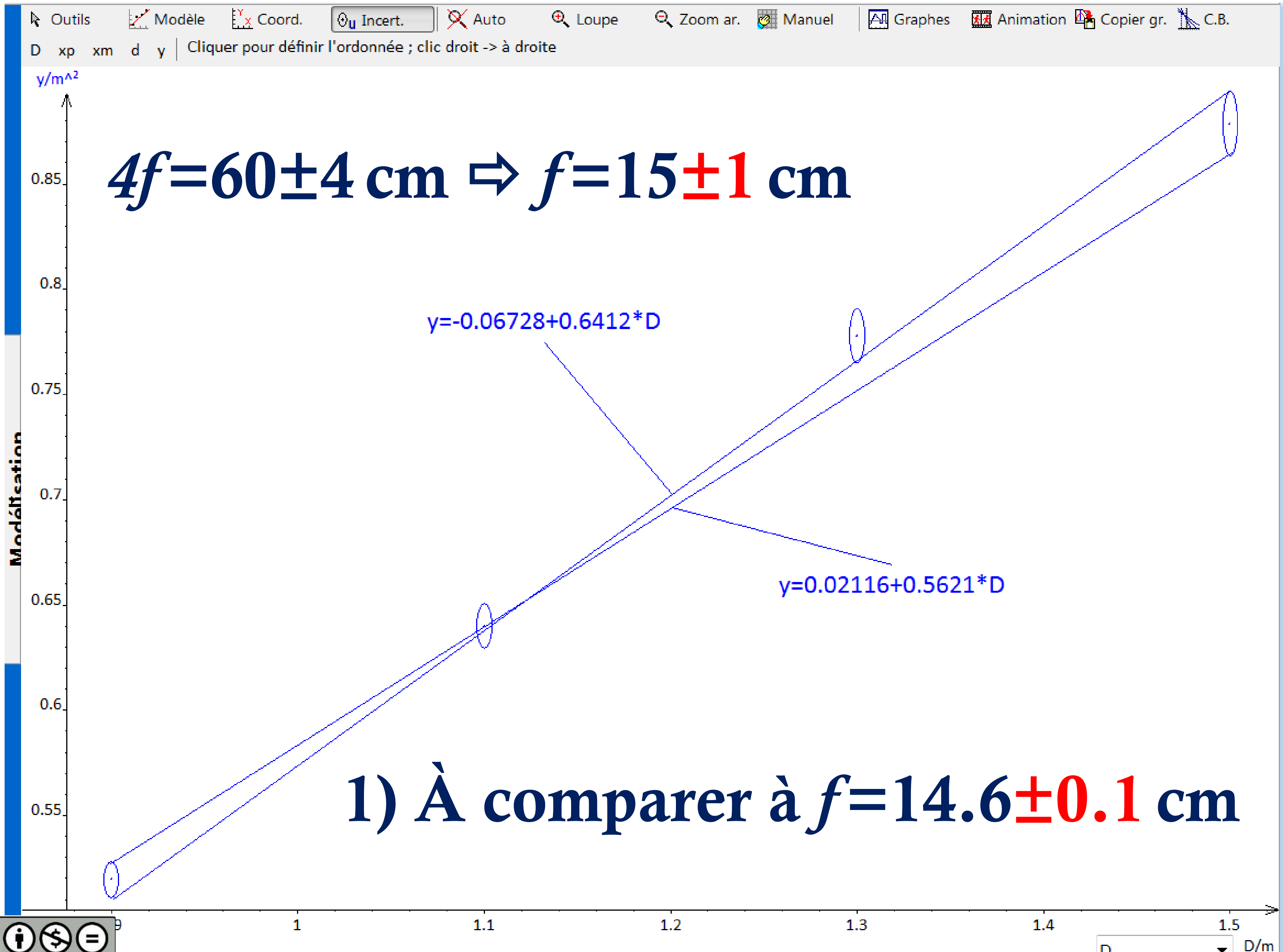
Utilisation de Regressi



$$\chi^2_{red}=1.13$$

$$f=14.6\pm0.1\ \mathrm{cm}$$

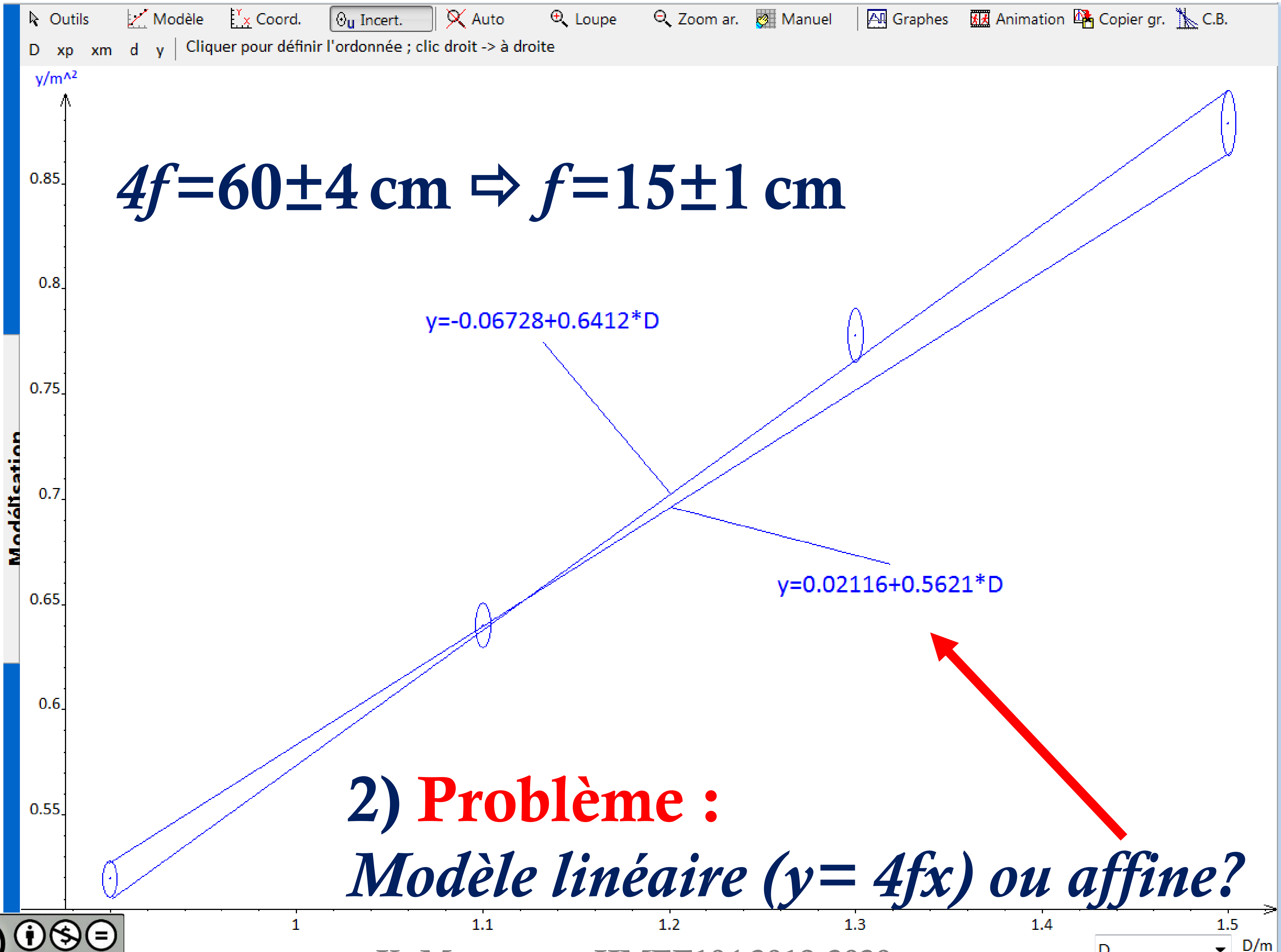
Méthode manuelle



$4f = 60 \pm 4$ cm ⇨ $f = 15 \pm 1$ cm

1) À comparer à $f = 14.6 \pm 0.1$ cm

Méthode manuelle



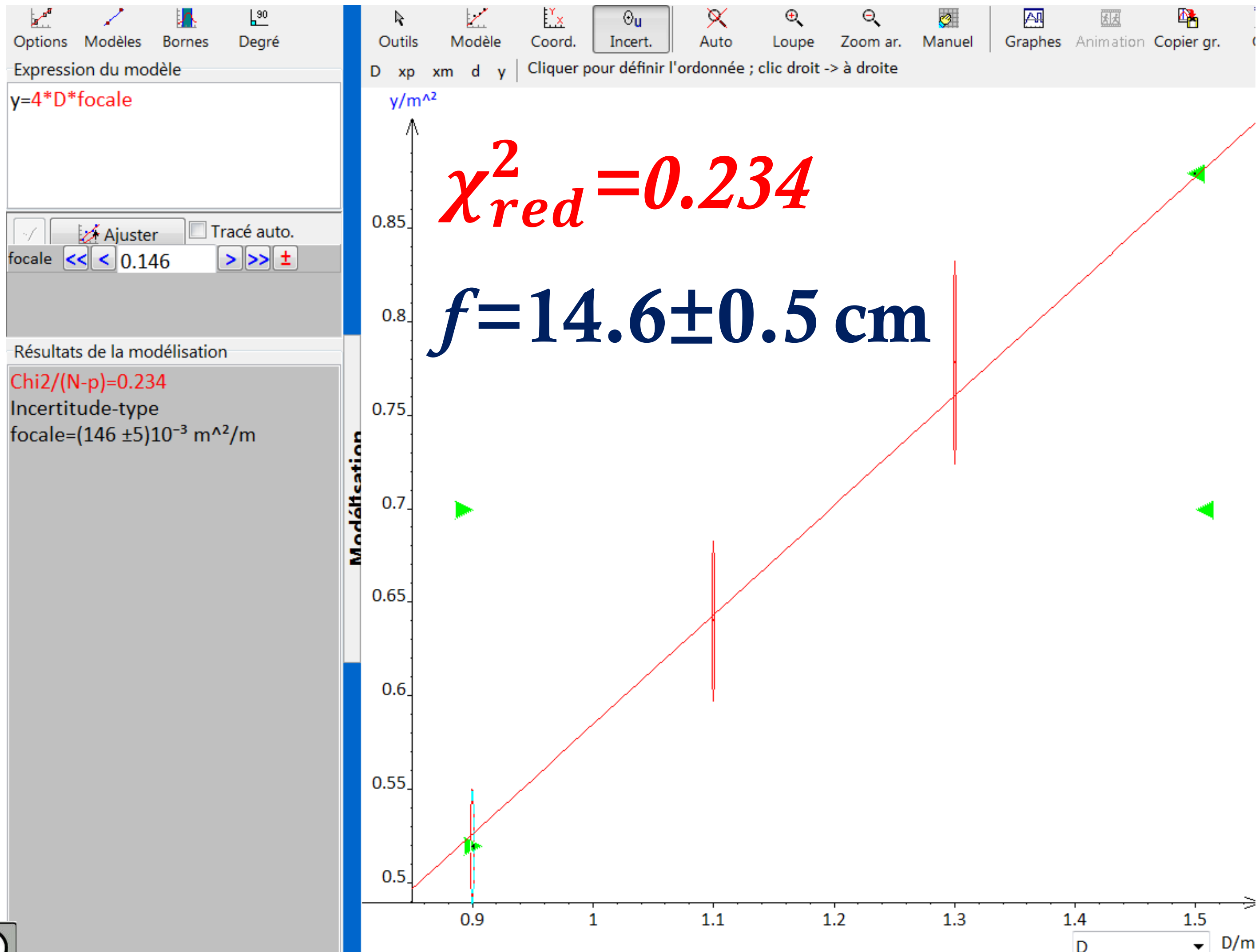
$4f = 60 \pm 4$ cm ⇨ $f = 15 \pm 1$ cm

2) Problème :

Modèle linéaire ($y = 4fx$) ou affine?

Utilisation de Regressi

Incertitudes sur-estimées



$$\chi^2_{red} = 0.234$$

$$f = 14.6 \pm 0.5\ \text{cm}$$

Utilisation de Regressi

Incertitudes TRES sur-estimées

Résultats de la modélisation

Chi2/(N-p)=1.13

Incertitude-type

focale=(146.2 ±1.0)10^{-3} m^2/m

Résultats de la modélisation

Chi2/(N-p)=0.0733

Incertitude-type

focale=(146 ±16)10^{-3} m^2/m

Résultats de la modélisation

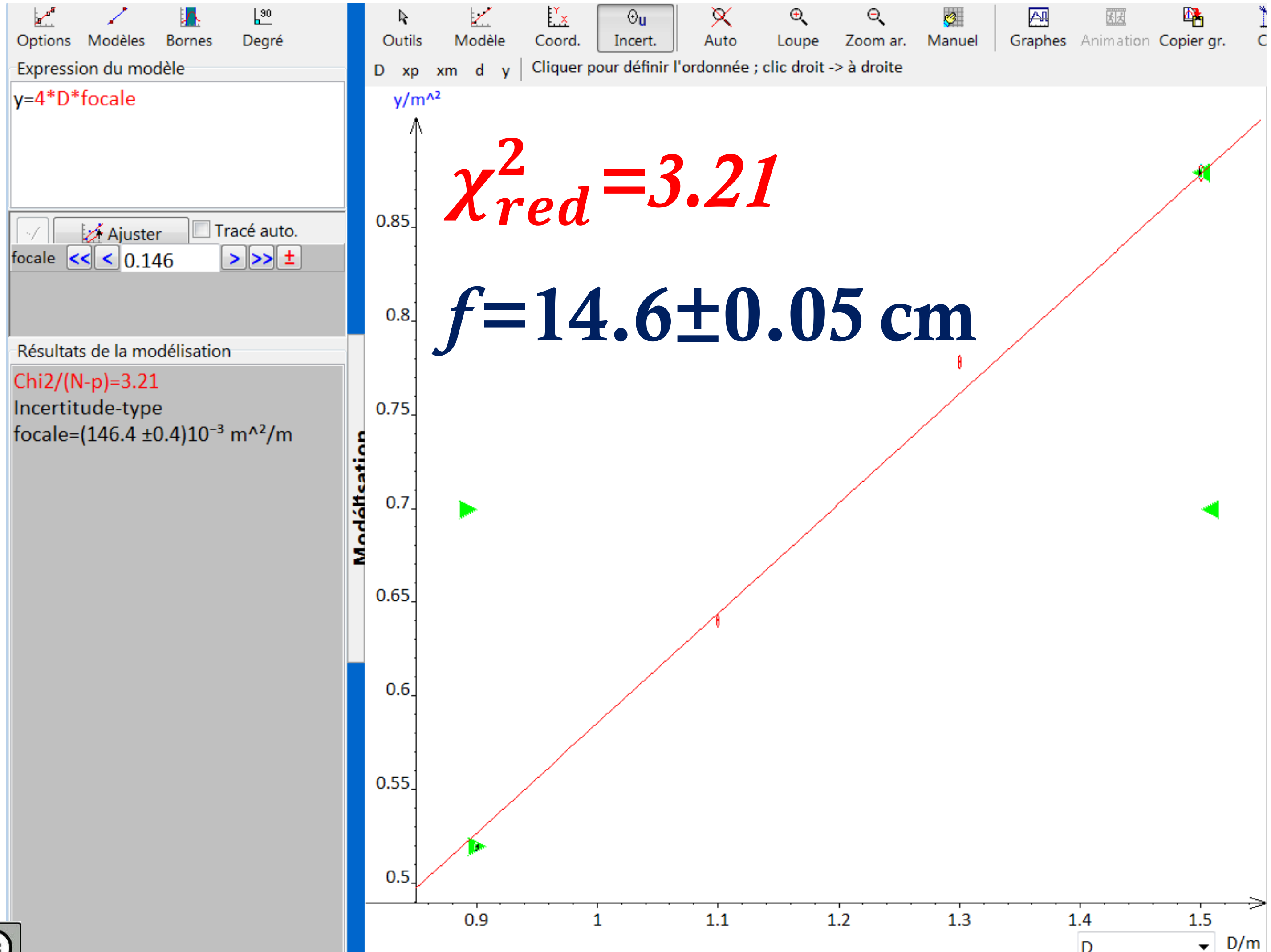
Chi2/(N-p)=0.0367

Incertitude-type

focale=(146 ±31)10^{-3} m^2/m

Utilisation de Regressi

Incertitudes sous-estimées



Utilisation de Regressi

Incertitudes TRES sous-estimées

Résultats de la modélisation

Chi2/(N-p)=1.13

Incertitude-type

focale=(146.2 ±1.0)10^{-3} m^2/m

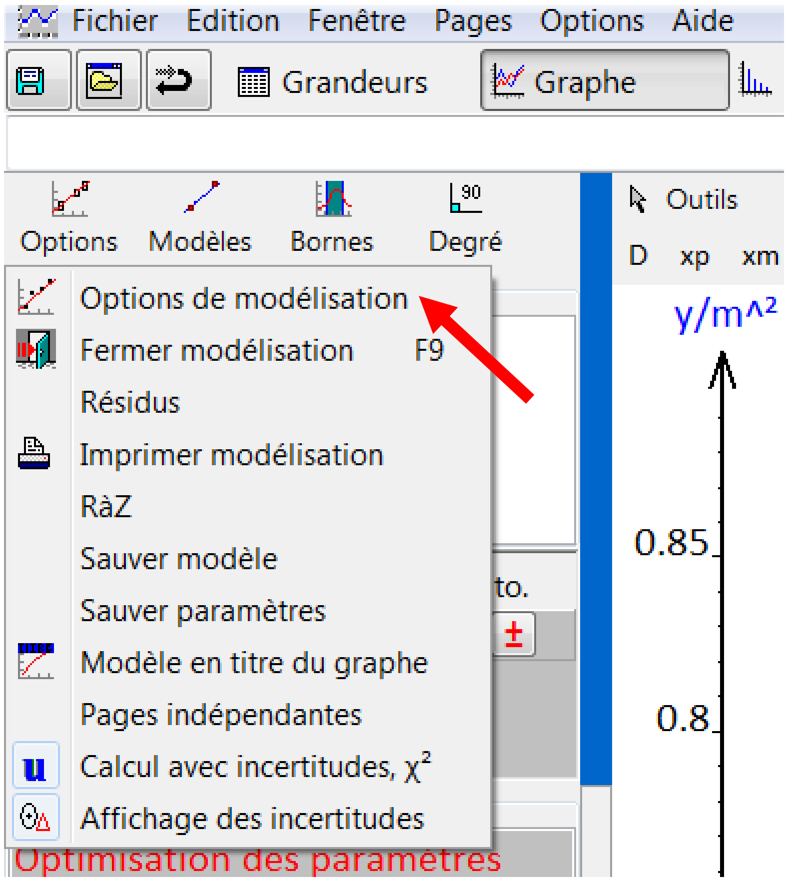
Résultats de la modélisation

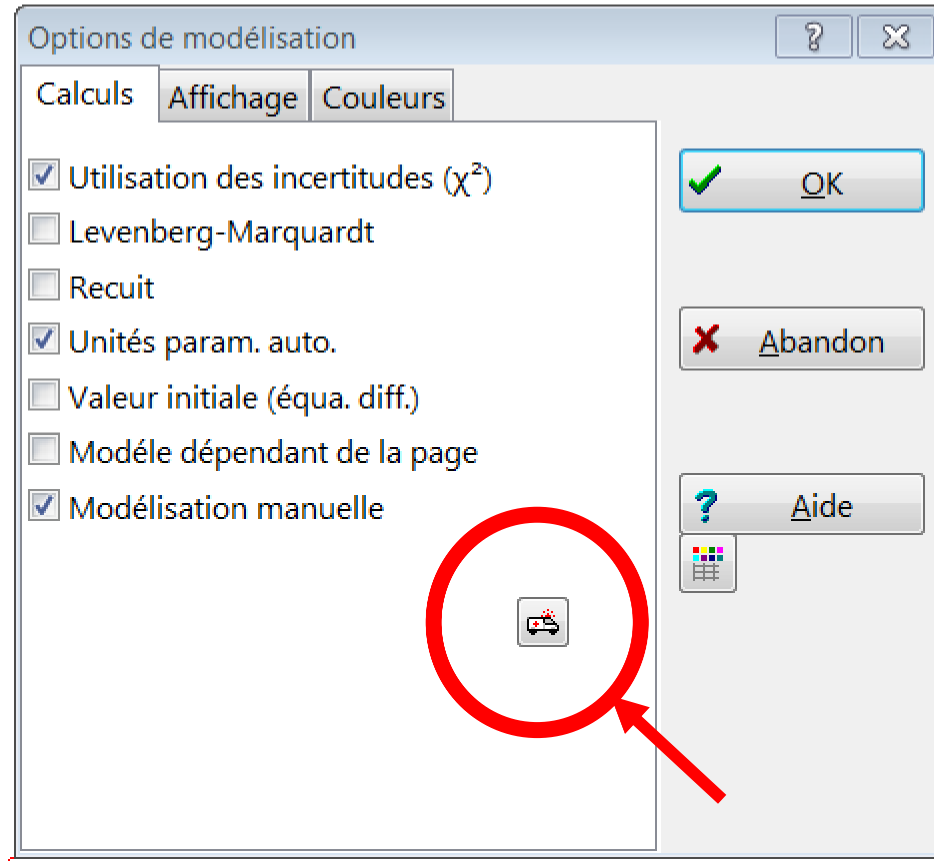
Chi2/(N-p)=5.36

Incertitude-type

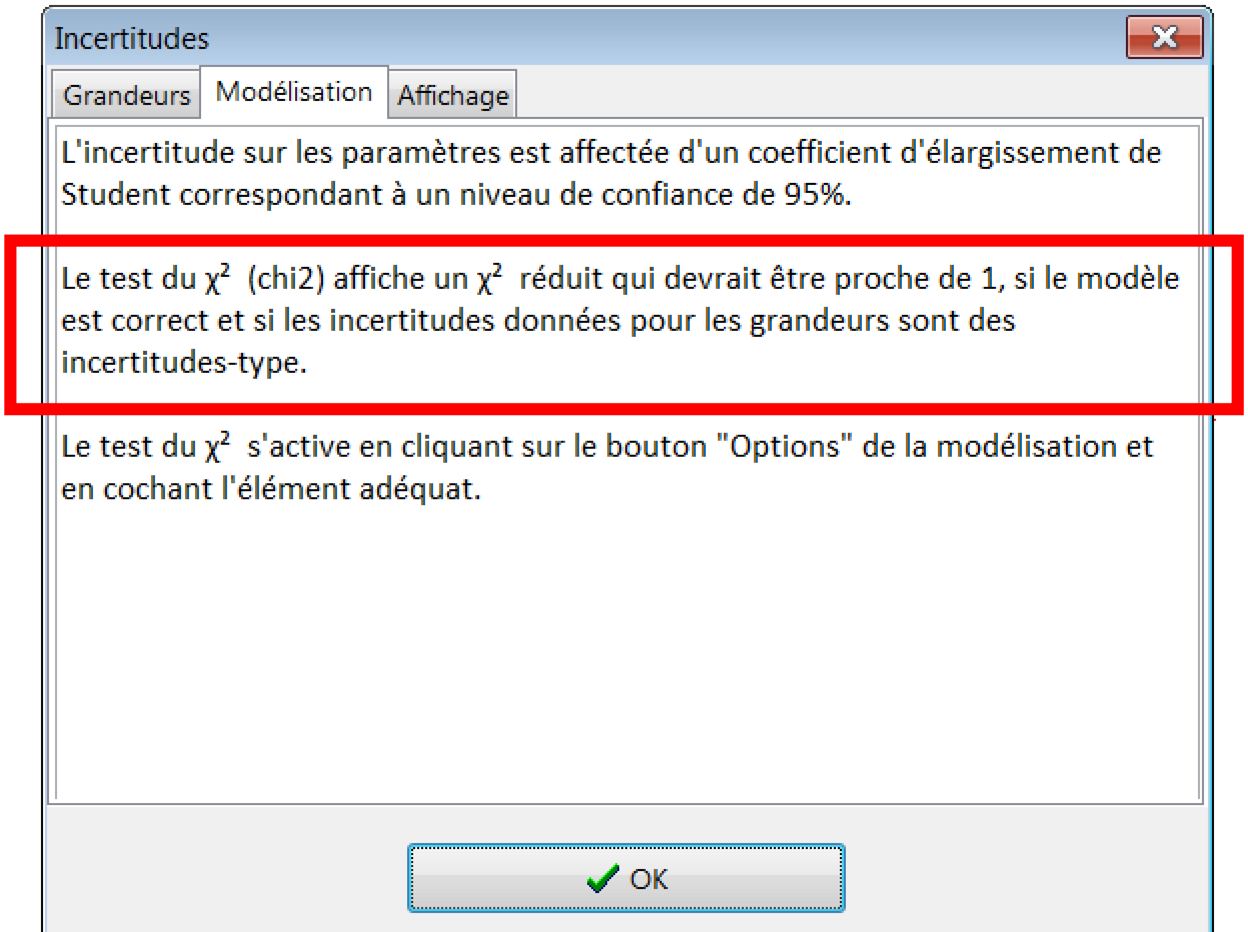
focale=(146.26 ±0.21)10^{-3} m^2/m

Utilisation de Regressi

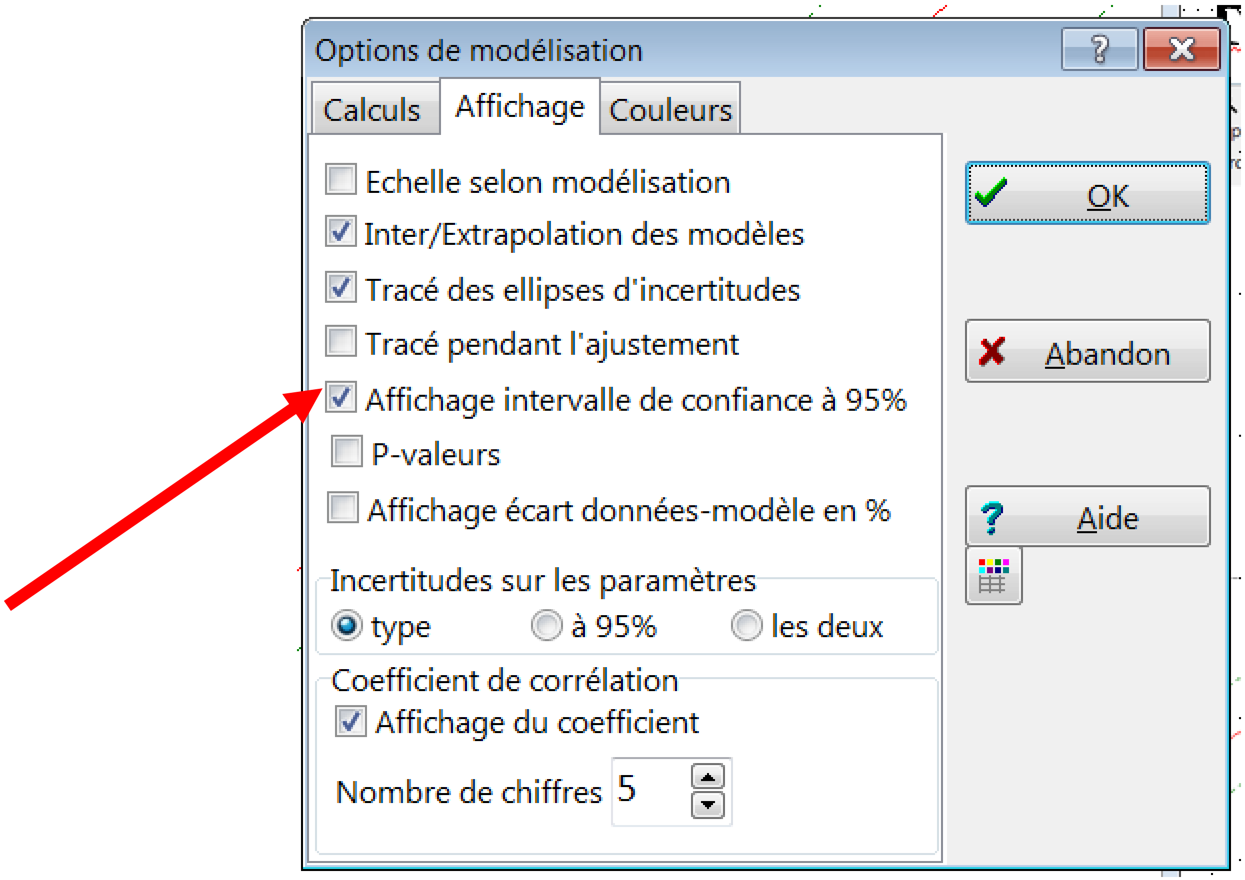




Utilisation de Regressi



Utilisation de Regressi



Utilisation de Regressi