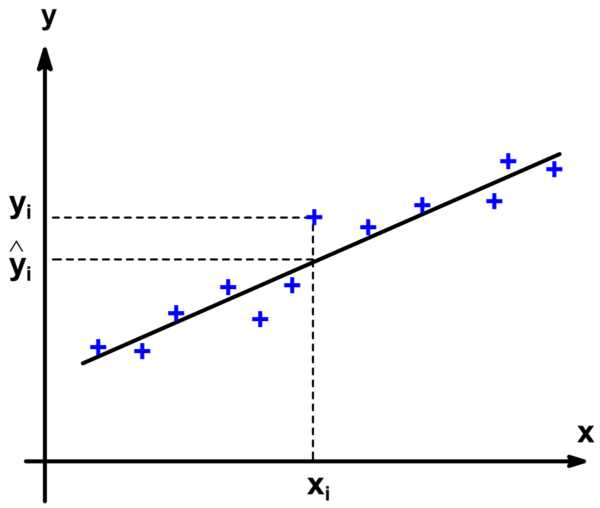
Régression linéaire

Conditions d'application :

- Absence d'erreur sur la variable x : les x_i doivent être connus de manière exacte. Une dispersion sur les x_i peut à la limite être acceptable si elle est négligeable par rapport à celle sur les y_i
- La dispersion sur les y_i doit être homogène et ne pas dépendre de la valeur des xi.



Soit $\hat{y}_i = \hat{a}_0 + \hat{a}_1 x_i$ l'équation de la droite optimale.

Cette droite est telle que la somme S des carrés des différences $(y_i - \hat{y}_i)$ soit minimale.

$$S = \sum_i \left[y_i - (\hat{a}_0 + \hat{a}_1 x_i)\right]^2$$

$$S = \sum_i \left[y_i^2 - 2 \times (\hat{a}_0 + \hat{a}_1 \times x_i) \times y_i + \hat{a}_1^2 x_i^2 + \hat{a}_0^2 + 2\hat{a}_0\hat{a}_1 x_i\right]$$

$$S = \sum_i y_i^2 - 2\hat{a}_1 \sum_i (y_i x_i) - 2\hat{a}_0 \sum_i (y_i) + \hat{a}_1^2 \sum_i x_i^2 + N\hat{a}_0^2 + 2\hat{a}_0\hat{a}_1 \sum_i x_i$$

$$\begin{cases} \dfrac{dS}{d\hat{a}_0} = -2\sum_i y_i + 2N\hat{a}_0 + 2\hat{a}_1 \sum_i x_i = 0 & (1) \\[2ex] \dfrac{dS}{d\hat{a}_1} = -2\sum_i (y_i x_i) + 2\hat{a}_1 \sum_i x_i^2 + 2\hat{a}_0 \sum_i x_i = 0 & (2) \end{cases}$$

On en tire :

$$\begin{cases} \hat{a}_0 N + \hat{a}_1 \sum_i x_i = \sum_i y_i & (1') \\ \hat{a}_0 \sum_i x_i + \hat{a}_1 \sum_i x_i^2 = \sum_i (y_i x_i) & (2') \end{cases}$$

et

$$\hat{a}_0 = \frac{\begin{vmatrix} \sum_i y_i & \sum_i x_i \\ \sum_i (y_i x_i) & \sum_i x_i^2 \end{vmatrix}}{\begin{vmatrix} N & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{vmatrix}} \qquad \hat{a}_1 = \frac{\begin{vmatrix} N & \sum_i y_i \\ \sum_i x_i & \sum_i x_i^2 \end{vmatrix}}{\begin{vmatrix} N & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{vmatrix}}$$

Soit :

$$\boxed{\hat{a}_0 = \frac{\sum_i y_i \times \sum_i x_i^2 - \sum_i x_i y_i \sum_i x_i}{N \sum_i x_i^2 - \left(\sum_i x_i\right)^2}} \qquad \boxed{\hat{a}_1 = \frac{N \sum_i (y_i x_i) - \sum_i x_i \sum_i y_i}{N \sum_i (x_i)^2 - \left(\sum_i x_i\right)^2}}$$

Ou encore :

$$\boxed{\begin{aligned} \hat{a}_0 &= \frac{\sum_i y_i - \hat{a}_1 \sum_i x_i}{N} \\ &= \bar{y} - \hat{a}_1 \bar{x} \end{aligned}}$$

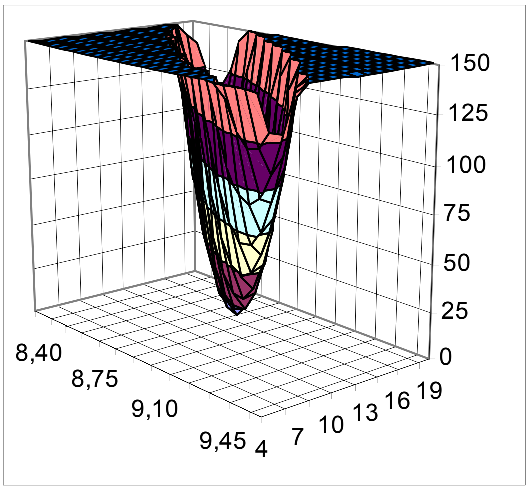
et

$$\boxed{\begin{aligned} \hat{a}_1 &= \frac{\sum_i (y_i x_i) - N\overline{xy}}{\sum_i (x_i)^2 - N\bar{x}^2} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} \\ &= \frac{\sum_i (x_i - \bar{x}) y_i}{\sum_i (x_i - \bar{x})^2} = \frac{\sum_i (y_i - \bar{y}) x_i}{\sum_i (x_i - \bar{x})^2} \\ &\text{puisque } \sum_i (y_i - \bar{y}) = \sum_i (x_i - \bar{x}) = 0 \end{aligned}}$$

avec $\bar{x} = \frac{\sum_i x_i}{N}$ et $\bar{y} = \frac{\sum_i y_i}{N}$.

Exercice RegLin1 : Etude de $S(a_0,a_1)$ pour une série de couples (x,y)

X	Y
2	30
4	48
6	65
8	85
10	101
12	122
14	140
16	155
18	175
20	191



$S=f(a_0,a_1)$

L'application des relation établies à la page précédente conduit à a_0 = 12,1 et a_1 = 9,0
Le tableau ci-dessous regroupe les valeurs de S pour différentes valeurs de a_0 et a_1 proches des valeurs optimales.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
18	a1	a0															
19	#N/A	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
20	8,40	2328	2042	1776	1530	1304	1098	912	746	600	474	368	282	216	170	144	138
21	8,45	2148	1873	1618	1383	1168	973	798	643	508	393	298	223	168	133	118	123
22	8,50	1977	1713	1469	1245	1041	857	693	549	425	321	237	173	129	105	101	117
23	8,55	1813	1560	1327	1114	921	748	595	462	349	256	183	130	97	84	91	118
24	8,60	1657	1415	1193	991	809	647	505	383	281	199	137	95	73	71	89	127
25	8,65	1509	1278	1067	876	705	554	423	312	221	150	99	68	57	66	95	144
26	8,70	1368	1148	948	768	608	468	348	248	168	108	68	48	48	68	108	168
27	8,75	1235	1026	837	668	519	390	281	192	123	74	45	36	47	78	129	200
28	8,80	1110	912	734	576	438	320	222	144	86	48	30	32	54	96	158	240
29	8,85	992	805	638	491	364	257	170	103	56	29	22	35	68	121	194	287
30	8,90	883	707	551	415	299	203	127	71	35	19	23	47	91	155	239	343
31	8,95	780	615	470	345	240	155	90	45	20	15	30	65	120	195	290	405
32	9,00	686	532	398	284	190	116	62	28	14	20	46	92	158	244	350	476
33	9,05	599	456	333	230	147	84	41	18	15	32	69	126	203	300	417	554
34	9,10	520	388	276	184	112	60	28	16	24	52	100	168	256	364	492	640
35	9,15	449	328	227	146	85	44	23	22	41	80	139	218	317	436	575	734
36	9,20	385	275	185	115	65	35	25	35	65	115	185	275	385	515	665	835
37	9,25	329	230	151	92	53	34	35	56	97	158	239	340	461	602	763	944
38	9,30	281	193	125	77	49	41	53	85	137	209	301	413	545	697	869	1061
39	9,35	240	163	106	69	52	55	78	121	184	267	370	493	636	799	982	1185
40	9,40	208	142	96	70	64	78	112	166	240	334	448	582	736	910	1104	1318
41	9,45	182	127	92	77	82	107	152	217	302	407	532	677	842	1027	1232	1457
42	9,50	165	121	97	93	109	145	201	277	373	489	625	781	957	1153	1369	1605
43	9,55	155	122	109	116	143	190	257	344	451	578	725	892	1079	1286	1513	1760
44	9,60	153	131	129	147	185	243	321	419	537	675	833	1011	1209	1427	1665	1923

Graphes 3D avec EXCEL

Pour tracer la courbe 3D ci-contre :

- Sélectionner les cellules $B$6:$Q$30
- Insertion→Graphique→Surface

Remarque : on ne dispose que **d'un seul axe numérique** (axe vertical). Pour représenter une fonction à trois dimensions, des valeurs peuvent être portées sur les deux autres axes mais elles doivent être **équidistantes.**
Exemple :
Sélectionner le graphique puis aller dans le sous menu «Données source» du menu «Graphique». Affecter à :

Nom : Pour la série 1, le contenu de la cellule $B$5
Pour la série 2, le contenu de la cellule $C$5
………
Pour la série 17, le contenu de la cellule $R$5.

Etiquettes des abscisses : $A$6:$A$30

Pour orienter la graphique, sélectionner le graphique puis Graphique → Mode 3D

Coefficient de corrélation

On montre que : $\sum_{i=1}^{N}(y_i-\overline{y})^2 = \sum_{i=1}^{N}(y_i-\hat{y}_i)^2 + \sum_{i=1}^{N}(\hat{y}_i-\overline{y})^2$

$\sum_{i=1}^{N}(y_i-\overline{y})^2$: variation totale

$\sum_{i=1}^{N}(y_i-\hat{y}_i)^2$: variation inexpliquée

$\sum_{i=1}^{N}(\hat{y}_i-\overline{y})^2$: variation expliquée

Coefficient de détermination :

$$r^2 = \frac{\sum_{i=1}^{N}(\hat{y}_i-\overline{y})^2}{\sum_{i=1}^{N}(y_i-\overline{y})^2} = \frac{\text{variation expliquée}}{\text{variation totale}}$$

r est le **coefficient de corrélation**.

Autre expression du coefficient de détermination :

$$\begin{aligned}
\sum_i [\hat{y}_i - y_i]^2 &= \sum_i [(\hat{a}_1 \times x_i + \hat{a}_0) - y_i]^2 \\
&= \sum_i \left[\hat{a}_1^2 x_i^2 + \hat{a}_0^2 + 2\hat{a}_1\hat{a}_0 x_i - \overline{y}^2\right] \\
&= \hat{a}_1^2 \sum_{i=1}^{N} x_i^2 + N\hat{a}_0^2 + 2\hat{a}_1\hat{a}_0 \sum_{i=1}^{N} x_i - N\overline{y}^2 \\
&= \hat{a}_1^2 \sum_{i=1}^{N} x_i^2 + N(\overline{y}-\hat{a}_1\overline{x})^2 + 2\hat{a}_1(\overline{y}-\hat{a}_1\overline{x})\sum_{i=1}^{N} x_i - N\overline{y}^2 \\
&= \hat{a}_1^2 \sum_{i=1}^{N} x_i^2 + N\overline{y}^2 + N\hat{a}_1^2\overline{x}^2 - 2N\hat{a}_1\overline{xy} + 2N\hat{a}_1\overline{xy} - 2N\hat{a}_1^2\overline{x}^2 - N\overline{y}^2 \\
&= \hat{a}_1^2 \left[\sum_{i=1}^{N} x_i^2 - N\overline{x}^2\right] = \hat{a}_1^2 \sum_i [x_i - \overline{x}]^2
\end{aligned}$$

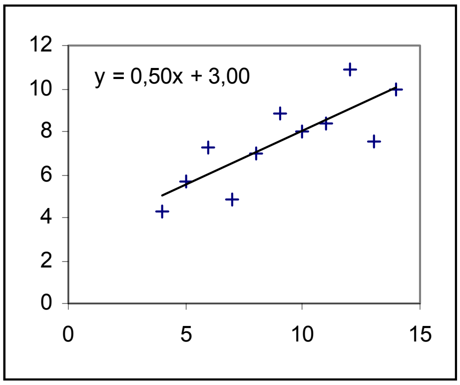
On en déduit :

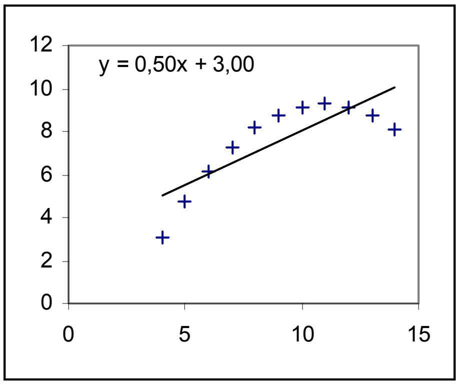
$$r^2 = \frac{\hat{a}_1^2 \sum_{i=1}^{N}(x_i - \bar{x})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} = \frac{\left[\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})\right]^2}{\left[\sum_{i=1}^{N}(x_i - \bar{x})^2\right]^2} \times \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} = \frac{\left[\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})\right]^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2 \sum_{i=1}^{N}(x_i - \bar{x})^2}$$

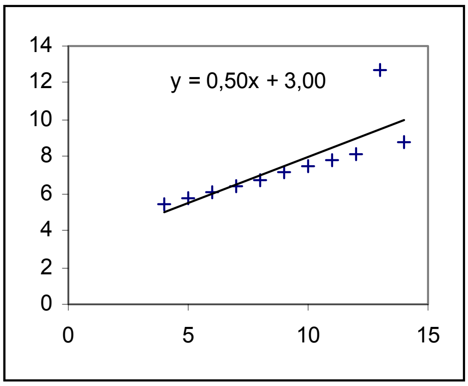
Dès que r^2 n'est pas strictement égal à un, le coefficient de corrélation ne peut plus être considéré comme un indicateur de linéarité.

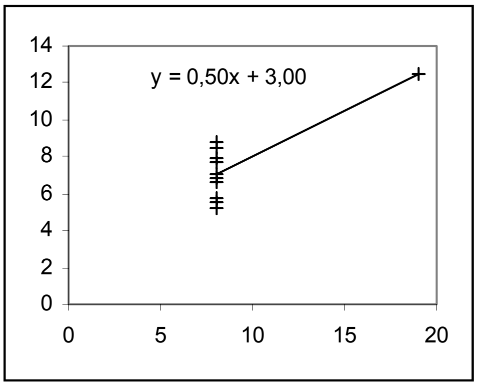
Les quatre 'nuages' de points ci-dessous ont mêmes moyennes, mêmes variances et même coefficient de corrélation (r = 0,82).

$D=r^2=$	0,67
$\sigma y=$	1,24
$\bar{x}=$	9,0
$\bar{y}=$	7,5









Variance résiduelle :

$$\sigma_y^2 = \frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{N-2}$$

« Explication » de la présence de N-2 au dénominateur : si on a seulement deux points, $\sum_{i=1}^{N} (\hat{y}_i - \overline{y}) = 0$ (la droite passe par les deux points). Pourtant les valeurs de y_i sont entachées d'incertitudes ! L'expression de σ_y^2 ci-dessus conduit alors à une valeur indéterminée : c'est le dénominateur N-2 qui permet mathématiquement cette indétermination (BUP 752 p 360).

Pour une démonstration de la relation ci-dessus, on pourra consulter :

YODALAH DODGE Analyse de régression appliquée Editions Dunod 1999 pages 28 à 30.

Régression linéaire et calcul matriciel :

Le système d'équations obtenu par minimisation du critère quadratique :

$$\begin{cases} N\hat{a}_0 + \hat{a}_1 \sum_i x_i = \sum_i y_i \quad (2') \\ \hat{a}_0 \sum_i x_i + \hat{a}_1 \sum_i x_i^2 = \sum_i (y_i x_i) \qquad (1') \end{cases}$$

peut s'écrire :

$$\begin{vmatrix} N & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{vmatrix} \times \begin{vmatrix} \hat{a}_0 \\ \hat{a}_1 \end{vmatrix} = \begin{vmatrix} \sum_i y_i \\ \sum_i (y_i x_i) \end{vmatrix}$$

dont on déduit :

$$\begin{vmatrix} \hat{a}_0 \\ \hat{a}_1 \end{vmatrix} = \begin{vmatrix} N & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{vmatrix}^{-1} \times \begin{vmatrix} \sum_i y_i \\ \sum_i (y_i x_i) \end{vmatrix}$$

Notons A la matrice $\begin{vmatrix} N & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{vmatrix}$

$$A^{-1} = \begin{vmatrix} N & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{vmatrix}^{-1} = \frac{1}{\det A} \begin{vmatrix} \sum_i x_i^2 & -\sum_i x_i \\ -\sum_i x_i & N \end{vmatrix}$$

avec

$$\det A = D = N\sum_i (x_i)^2 - \left(\sum_i x_i\right)^2 = N\sum_i (x_i)^2 - (N\bar{x})^2 = N\left[\sum_i (x_i - \bar{x})^2\right]$$

Variance et covariance des variables aléatoires $\hat{a}_1$ et $\hat{a}_0$

Les démonstration des relations ci-dessous sont données par exemple dans : YODALAH DODGE Analyse de régression appliquée Editions Dunod 1999 page 27.

$$\mathrm{var}(\hat{a}_0) = \sigma^2_{\hat{a}_0} = \sigma^2_y \frac{\sum_{i=1}^{N} x_i^2}{D} = \sigma^2_y \frac{\sum_{i=1}^{N} x_i^2}{N\sum_{i=1}^{N}(x_i - \overline{x})^2}$$

$$\mathrm{var}(\hat{a}_1) = \sigma^2_{\hat{a}_1} = \sigma^2_y \frac{N}{D} = \sigma^2_y \frac{N}{N\sum_i (x_i - \overline{x})^2} = \frac{\sigma^2_y}{\sum_i (x_i - \overline{x})^2}$$

$$\mathrm{cov}(\hat{a}_0,\hat{a}_1) = -\sigma^2_y \frac{\sum_{i=1}^{N} x_i}{D} = -\sigma^2_y \frac{\sum_{i=1}^{N} x_i}{N\sum_{i=1}^{N}(x_i - \overline{x})^2} \text{ avec } \sigma^2_y = \frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N-2}$$

On définit la **matrice de variance-covariance** V par :

$$V = \begin{vmatrix} \mathrm{var}(\hat{a}_0) & \mathrm{cov}(\hat{a}_0,\hat{a}_1) \\ \mathrm{cov}(\hat{a}_0,\hat{a}_1) & \mathrm{var}(\hat{a}_0) \end{vmatrix}$$

On vérifiera facilement qu'elle est égale à $\sigma^2_y \times |A|^{-1}$

<u>Estimation des incertitudes sur les valeurs de</u> $\hat{a}_1$ et $\hat{a}_0$

Les quantités $\frac{a_1 - \hat{a}_1}{\sigma_{\hat{a}_1}}$ et $\frac{a_0 - \hat{a}_0}{\sigma_{\hat{a}_0}}$ suivent une loi de Student à N-2 degrés de liberté (a_0 et a_1 y désignent les « vraies valeurs » des paramètres).

Les intervalles de confiance (à un niveau de confiance donné) sont donc donnés par les relations suivantes :

$$\hat{a}_0 - t_{N-2,\alpha} \times \sigma_{\hat{a}_0} \text{ , } \hat{a}_0 + t_{N-2,\alpha} \times \sigma_{\hat{a}_0}$$

$$\hat{a}_1 - t_{N-2,\alpha} \times \sigma_{\hat{a}_1} \text{ , } \hat{a}_1 + t_{N-2,\alpha} \times \sigma_{\hat{a}_1}$$

ou $t_{N-2,\alpha}$ désigne le coefficient le coefficient de Student à (N-2) degrés de liberté au taux de confiance α choisi.

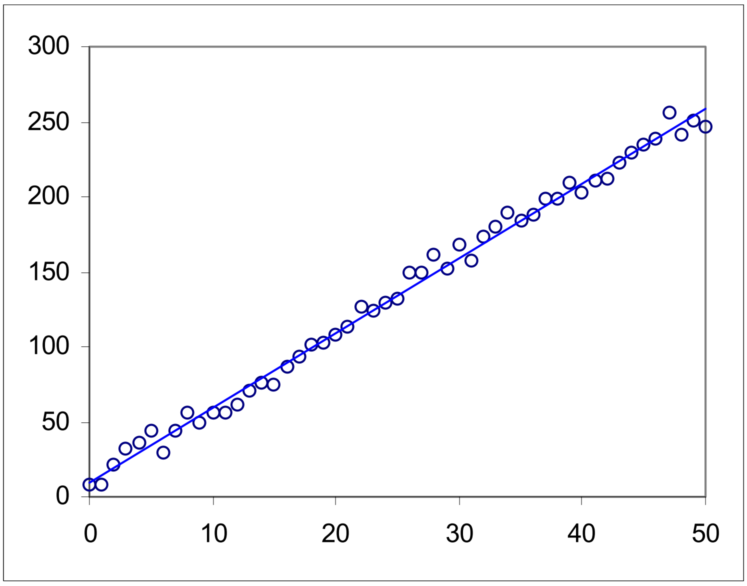
Exercice : RegLin2

Dans cet exercice, on se propose de générer des données bruitées y=a0+a1*x+bruit. On modélise ensuite les données bruitées par $a0_{MC}+a1_{MC}*x$. On compare alors les valeurs de $a0_{MC}$ et a0 d'une part et $a1_{MC}$ et a1 d'autre part. ($a0_{MC}$ et $a1_{MC}$ désignent les estimateurs de a0 et a1 obtenus par la méthode des moindres carrés)

	A	B	C	D	E	F	G	H
1	**a0**	**a1**	**bmax**	**a0MC**	**a1MC**	**sigmay**	**sigmaa0**	**sigmaa1**
2	10	5	5					
3								
4	**Sx**	**Sy**	**N**	**Sxy**	**Sxx**	**denom**		**SCR**
5								
6								
7	**x**	**y**	**g**	**xy**	**xx**	**ycalc**	**R**	**RR**
8								
9	0							
10	1							
	2							
	3							
	4							
	5							
59	50							

1) Ecrire les titres des variables dans les cellules
2) Entrer les valeurs de a0, a1 et bmax dans les cellules A2, B2 et C2
3) Affecter les noms a0, a1 et bmax aux variables figurants respectivement dans les cellules A2, B2 et C2. (**Insertion – Nom – Créer**)
4) Créer la suite des valeurs de x ($0 \leq x \leq 50$) dans ls cellules A9 à A35. (On écrit les 2 premières lignes puis on recopie vers le bas).
5) Créer la série de valeurs de g : **OUTILS** puis **Utilitaire d'analyse** puis **Génération de nombre aléatoires**. Dans la boîte de dialogue choisir :
   Distribution : Normale (valider)

   1. **Paramètres : Moyenne = 0 ; écart type = 1**
      a) Plage de sortie : C9 :C35
6) Ecrire la formule =a0+a1*A9+bmax*C9 dans la cellule B9.
   Recopier la formule dans les cellules B10 à B35 pour générer les valeurs bruitées de y.



7) Ecrire la formule =A9*B9 dans la cellule D9 et la formule = A9*A9 dans la cellule E9. Recopier respectivement ces formule dans les cellules D10 à D35 et E10 à E35.
8) Pour calculer les paramètre de la régression linéaire (MC) : a0MC , a1MC, …..il faut connaître les valeurs de $\sum_1^N x_i$, $\sum_1^N y_i$, $\sum_1^N x_i y_i$,…

   Ecrire les formules :
   - =somme(A9 :A35) dans la cellule A5
   - =somme(B9 :B35) dans la cellule B5
   - =NBVAL(A9 :A100) dans la cellule C5
   - =somme(D9 :D35) dans la cellule D5
   - =somme(E9 :E35) dans la cellule E5
   - =C5*E5-A5^2 dans la cellule F5

9) Les valeurs de $a0_{MC}$ et $a1_{MC}$ peuvent alors être calculées à l'aide des formules correspondantes.

10) Déterminer alors les valeurs de ycalc = $a0_{MC}$+$a1_{MC}$*x en écrivant = $D$1+$E$1*A9 dans la cellule F9 puis en recopiant jusqu'à F35.

11) Calculer alors les valeurs des résidus R = ycalc-yi, RR= R*R puis SCR (Somme de Carrés des Résidus) dans les cellules prévues.

12) Déterminer alors les valeurs de σ(y) ; σ(a0) ; σ(a1)

13) Conclusion.

14) EXCEL est capable de calculer directement les principaux paramètres de la régression linéaire. Ces valeurs sont inscrites dans une plage de 10 cellules (5 lignes de 2 colonnes).

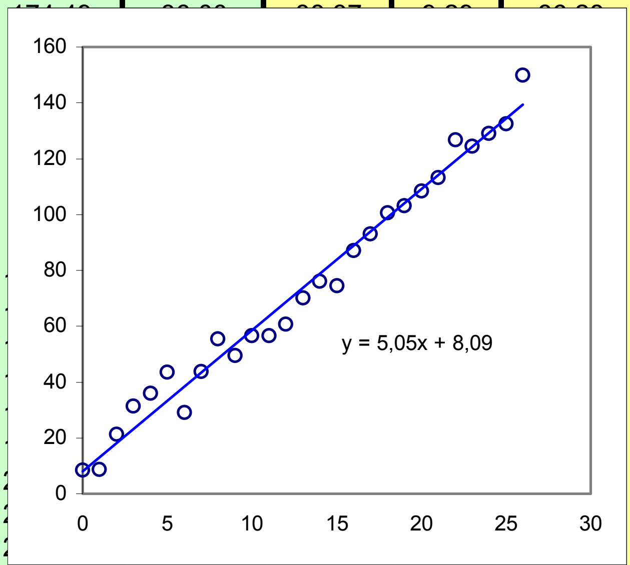
Sélectionner une plage de 10 cellules (5 lignes et 2 colonnes). Puis écrire =DROITEREG(B9 :B35 ;A9 :A35 ;VRAI ;VRAI) dans la barre de formules. Valider en appuyant simultanément sur les touches CTRL SHIFT ENTER.

Les statistiques de régression sont les suivantes :

a1	a0
σ(a1)	σ(a0)
D (coefficient de détermination)	σ(y)
	N-2
	SCR

15) Utiliser les valeurs de σ(a0) et σ(a1) pour déterminer les incertitudes sur les valeurs de a_0 et a_1 pour un taux de confiance de 95%.

	A	B	C	D	E	F	G	H
1	**a0**	**a1**	**bmax**	**a0MC**	**a1MC**	**σ(y)**	**σ(a0)**	**σ(a1)**
2	10	5	5	8,09	5,05	5,72	2,14	1,41E-01
3								
4	**Sx**	**Sy**	**N**	**Sxy**	**Sxx**	**denom**		**SSR**
5	351	1989,99	27,00	34137,70	6201	44226		818,87
6								
7	**x**	**y**	**g**	**xy**	**xx**	**ycalc**	**R**	**RR**
8								
9	0	8,50	-0,30	0,00	0,00	8,09	-0,41	0,17
10	1	8,61	-1,28	8,61	1,00	13,13	4,52	20,45
11	2	21,22	0,24	42,44	4,00	18,18	-3,04	9,24
12	3	31,38	1,28	94,15	9,00	23,23	-8,15	66,49
13	4	35,99	1,20	143,97	16,00	28,28	-7,72	59,53
14	5	43,67	1,73	218,33	25,00	33,32	-10,34	106,96
15	6	29,08	-2,18					
16	7	43,83	-0,23					
17	8	55,48	1,10					
18	9	49,57	-1,09					
19	10	56,55	-0,69					
20	11	56,55	-1,69					
21	12	60,77	-1,85					
22	13	70,11	-0,98					
23	14	76,13	-0,77					
24	15	74,41	-2,12					
25	16	87,16	-0,57					
26	17	92,98	-0,40					
27	18	100,67	0,13					
28	19	103,17	-0,37					
29	20	108,37	-0,33					
30	21	113,15	-0,37					
31	22	126,71	1,34					
32	23	124,57	-0,09	2865,19	529,00	124,18	-0,40	0,16
33	24	129,07	-0,19	3097,66	576,00	129,23	0,16	0,02
34	25	132,43	-0,51	3310,85	625,00	134,27	1,84	3,38
35	26	149,86	1,97	3896,39	676,00	139,32	-10,54	111,10
36								



<u>Avec REGRESSI (REGWIN) (extrait de la notice)</u>

Les valeurs des paramètres sont données sous la forme A±ΔA. L'intervalle [A-ΔA, A+ΔA] correspond à un intervalle de confiance de Student à 95 %. La précision relative sur la fonction est le rapport entre la moyenne quadratique des écarts fonction théorique/expérimentale et la moyenne quadratique de la fonction expérimentale. Lorsque l'intervalle de confiance est plus grand ou de l'ordre de grandeur de la valeur du paramètre, le logiciel signale que ce paramètre n'est pas significatif en remplaçant la précision par un ?. De même, lorsque le logiciel trouve une précision relative trop importante, pour signaler que la fonction proposée est mal adaptée, le logiciel remplace la précision par un ?.
Si vous avez défini des incertitudes, l'ajustement se fait selon la méthode de l'ellipse et on donne la valeur de $\chi^2/(N-r)$ avec N nombre de points et r nombre de paramètres, cette valeur devrait être proche de 1 : une valeur plus grande signifie que la fonction est inadaptée, plus petite que cela marche trop bien !

Ecart quad. y=3.681 = $\sqrt{13,552}$

$a_1 = 9.01 \pm 0.16$ (0,16 = 0,072 * 2,306 = 0,166)

$a_0 = 2.1 \pm 2.0$ (2,0 = 0,889 * 2,306 = 2,050)

S = 13,552

Incertitude sur y $= \sigma_y \times t_{0,05,N-2}$

$$\text{Ecart relatif (en\%)} = \frac{\sqrt{\sum_{i=1}^{Nexp}\left[y_i - (a_1 x_i + a_0)\right]^2}}{\sqrt{\sum_{i=1}^{Nexp}\left[y_i\right]^2}} = \frac{\sqrt{13,552}}{\sqrt{150470}} = 0,95\%$$

<u>Exercice RegLin3 :</u> Traiter les données de la page suivante avec **Regressi**

	B	C	D	E	F	G	H	I	J
2	**x**		**y**		**xy**		**x^2**		**y^2**
3	2		30		60		4		900
4	4		48		192		16		2304
5	6		65		390		36		4225
6	8		85		680		64		7225
7	10		101		1010		100		10201
8	12		122		1464		144		14884
9	14		140		1960		196		19600
10	16		155		2480		256		24025
11	18		175		3150		324		30625
12	20		191		3820		400		36481
13	**110**	**Sy=**	**1112**	**Sxy=**	**15206**	**Sx_2=**	**1540**	**Sy_2=**	**150470**

Utilisation des formules :

17	(10*Sxy-Sx*Sy)/(10*Sx_2-(Sx)^2)	=	**a1**	**9,01**
19	(Sx_2*Sy-Sxy*Sx)/(10*Sx_2-(Sx)^2)	=	**a0**	**12,07**
21	(1/(10-2)*(Sy_2-p*Sxy-q*Sy))^0,5	=	**σ_y**	**1,30**
23	s_y*(10/(10*Sx_2-(Sx)^2))^0,5	=	**σa1**	**0,072**
25	s_y*(Sx_2/(10*Sx_2-Sx^2))^0,5	=	**σa0**	**0,89**

28	INDEX(DROITEREG(D3:D12;B3:B12;VRAI;VRAI);1;1)=	9,01
29	INDEX(DROITEREG(D3:D12;B3:B12;VRAI;VRAI);1;2)=	12,07
30	INDEX(DROITEREG(D3:D12;B3:B12;VRAI;VRA));2;1)=	0,072
31	INDEX(DROITEREG(D3:D12;B3:B12;VRAI;VRAI);2;2)=	0,89

=DROITEREG(D3:D12;B3:B12;vrai;vrai)

9,012	**12,067**
0,072	**0,889**
0,999	**1,302**
15822,3	**8,000**
26802,0	**13,552**

Formules matricielles et modalités de saisie
Sélectionnez la ou les cellules qui vont contenir la formule, créez la formule et appuyez sur CTRL+MAJ+ENTRÉE pour taper la formule.

1) Sélectionner une plage de 5 lignes et 2 colonnes
2) taper la formule DROITEREG(……)
3) Valider avec CTRL+MAJ+ENTREE

Variance d'une fonction F des paramètres a_0 et a_1

Les variances $\mathrm{var}(\hat{a}_0)$, $\mathrm{var}(\hat{a}_1)$ et la covariance $\mathrm{cov}(\hat{a}_0,\hat{a}_1)$ sont connues (Cf. page 9). La variance de la fonction $F(\hat{a}_0,\hat{a}_1)$ des variables aléatoires $\hat{a}_0$ et $\hat{a}_1$ est donnée par la relation suivante (loi de propagation des erreurs) :

$$\sigma_F^2 \approx \left(\frac{\partial F}{\partial \hat{a}_0}\right)^2 \sigma_{\hat{a}_0}^2 + \left(\frac{\partial F}{\partial \hat{a}_1}\right)^2 \sigma_{\hat{a}_1}^2 + 2\left(\frac{\partial F}{\partial \hat{a}_0}\right)\left(\frac{\partial F}{\partial \hat{a}_1}\right)\mathrm{cov}(\hat{a}_0,\hat{a}_1)$$

que l'on peut écrire sous forme matricielle :

$$\sigma_F^2 = \begin{vmatrix} \frac{\partial F}{\partial a_0} & \frac{\partial F}{\partial a_1} \end{vmatrix} \times \begin{vmatrix} \mathrm{var}(\hat{a}_0) & \mathrm{cov}(\hat{a}_0,\hat{a}_1) \\ \mathrm{cov}(\hat{a}_0,\hat{a}_1) & \mathrm{var}(\hat{a}_1) \end{vmatrix} \times \begin{vmatrix} \frac{\partial F}{\partial a_0} \\ \frac{\partial F}{\partial a_1} \end{vmatrix}$$

soit :

$$\sigma_F^2 = \sigma_y^2 \begin{vmatrix} \frac{\partial F}{\partial a_0} & \frac{\partial F}{\partial a_1} \end{vmatrix} \times |A|^{-1} \times \begin{vmatrix} \frac{\partial F}{\partial a_0} \\ \frac{\partial F}{\partial a_1} \end{vmatrix} = \sigma_y^2 \frac{1}{D} \times \begin{vmatrix} \frac{\partial F}{\partial a_0} & \frac{\partial F}{\partial a_1} \end{vmatrix} \times \begin{vmatrix} \sum_{i=1}^{N} x_i^2 & -\sum_{i=1}^{N} x_i \\ -\sum_{i=1}^{N} x_i & N \end{vmatrix} \times \begin{vmatrix} \frac{\partial F}{\partial a_0} \\ \frac{\partial F}{\partial a_1} \end{vmatrix}$$

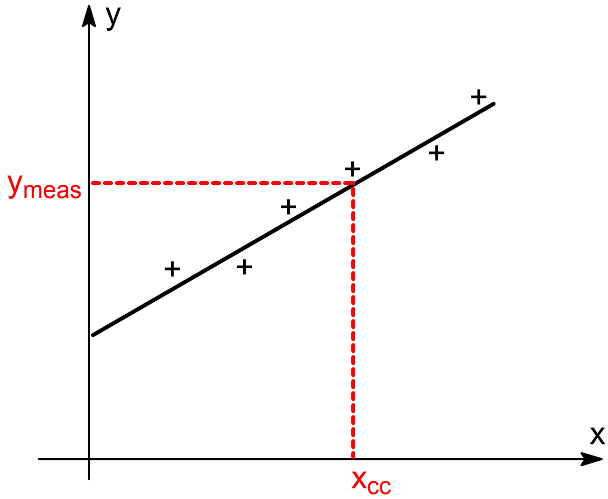
avec

$$A^{-1} = \begin{vmatrix} N & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{vmatrix}^{-1} = \frac{1}{\det A} \begin{vmatrix} \sum_i x_i^2 & -\sum_i x_i \\ -\sum_i x_i & N \end{vmatrix}$$

et $\det A = D = N\sum_i (x_i)^2 - \left(\sum_i x_i\right)^2$

- **calibration (dosage par étalonnage)**

On établit expérimentalement une relation entre y et x en mesurant les valeurs de y pour des valeurs connues de x. On mesure ensuite y_{meas} pour une valeur de x (x_{cc}) inconnue. On utilise la courbe d'étalonnage pour déterminer la valeur de x_{cc}.



La fonction F est la fonction qui à une valeur mesurée de y (y_{meas}) fait correspondre une valeur de x (x_{cc}) :

$$F = x_{cc} = \frac{y_{meas} - a_0}{a_1}$$

$$\frac{\partial x_{cc}}{\partial a_0} = -\frac{1}{a_1} \qquad \qquad \frac{\partial x_{cc}}{\partial a_1} = -\frac{y_{meas} - a_0}{a_1^2} = -\frac{x_{cc}}{a_1}$$

$$\sigma_F^2 = \sigma_y^2 \frac{1}{D} \times \begin{vmatrix} -\dfrac{1}{a_1} & -\dfrac{x_{cc}}{a_1} \end{vmatrix} \times \begin{vmatrix} \sum\limits_{i=1}^{N} x_i^2 & -\sum\limits_{i=1}^{N} x_i \\ -\sum\limits_{i=1}^{N} x_i & N \end{vmatrix} \times \begin{vmatrix} -\dfrac{1}{a_1} \\ -\dfrac{x_{cc}}{a_1} \end{vmatrix}$$

$$= \sigma_y^2 \frac{1}{D} \times \begin{vmatrix} -\dfrac{1}{a_1} \sum\limits_{i=1}^{N} x_i^2 + \dfrac{x_{cc}}{a_1} \sum\limits_{i=1}^{N} x_i & \dfrac{1}{a_1} \sum\limits_{i=1}^{N} x_i - \dfrac{N x_{cc}}{a_1} \end{vmatrix} \times \begin{vmatrix} -\dfrac{1}{a_1} \\ -\dfrac{x_{cc}}{a_1} \end{vmatrix}$$

$$= \sigma_y^2 \frac{1}{D} \times \left[-\frac{1}{a_1} \sum_{i=1}^{N} x_i^2 + \frac{x_{cc}}{a_1} \sum_{i=1}^{N} x_i \right] \times \left(-\frac{1}{a_1} \right) - \frac{x_{cc}}{a_1} \times \left[\frac{1}{a_1} \sum_{i=1}^{N} x_i - \frac{N x_{cc}}{a_1} \right]$$

$$\boxed{\sigma_F^2 = \sigma_y^2 \frac{1}{D} \times \frac{1}{a_1^2} \left[\sum_{i=1}^{N} x_i^2 - 2 x_{cc} \sum_{i=1}^{N} x_i + N x_{cc}^2 \right]}$$

La détermination de x_{cc} met en jeu la fonction F et la valeur y_{meas} .

La variance de F est connue ; $\sigma_y^2 = \frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N-2}$ est un bon estimateur de $\sigma_{y_{meas}}^2$.

La loi de propagation des erreurs s'écrit alors : $\sigma_{x_{cc}}^2 = \left(\frac{\partial x_{cc}}{\partial y_{meas}}\right)^2 \sigma_{y_{meas}}^2 + \sigma_F^2$

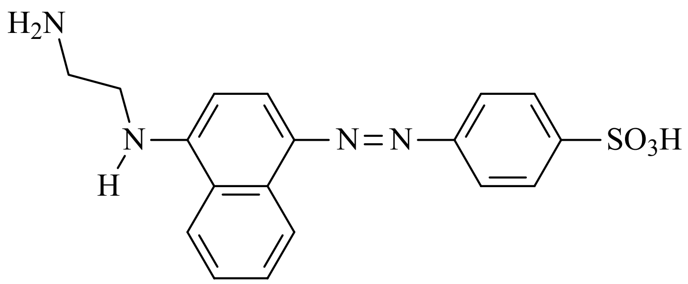
Soit, en tenant compte de l'expression de σ_F^2 :

$$\boxed{\sigma_{x_{cc}}^2 = \frac{\sigma_y^2}{a_1^2}\left[1 + \frac{1}{D}\left(Nx_{cc}^2 - 2x_{cc}\sum_{i=1}^{N} x_i + \sum_{i=1}^{N} x_i^2\right)\right]}$$

Remarque : $Nx_{cc}^2 - 2x_{cc}\sum_{i=1}^{N} x_i + \sum_{i=1}^{N} x_i^2$ est minimal si $x_{cc} = \frac{\sum_{i=1}^{N} x_i}{N}$

L'exemple suivant est relatif au dosage des ions nitrite dans un jambon.

Les ions nitrite sont utilisés pour préparer un colorant azoïque :

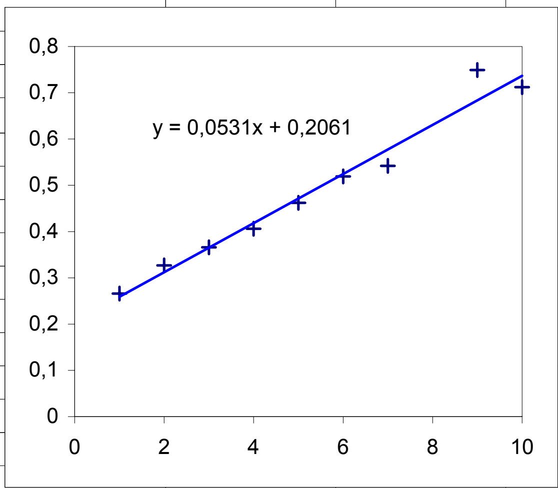


qui absorbe dans le visible.

On réalise la gamme d 'étalonnage en mesurant l'absorbance de solutions du colorant azoïque préparées à partir de solutions d'ions nitrite de concentrations connues puis on mesure de l'absorbance d'une solution dans laquelle les ions nitrite ont été apportés par un morceau de jambon de masse connue.

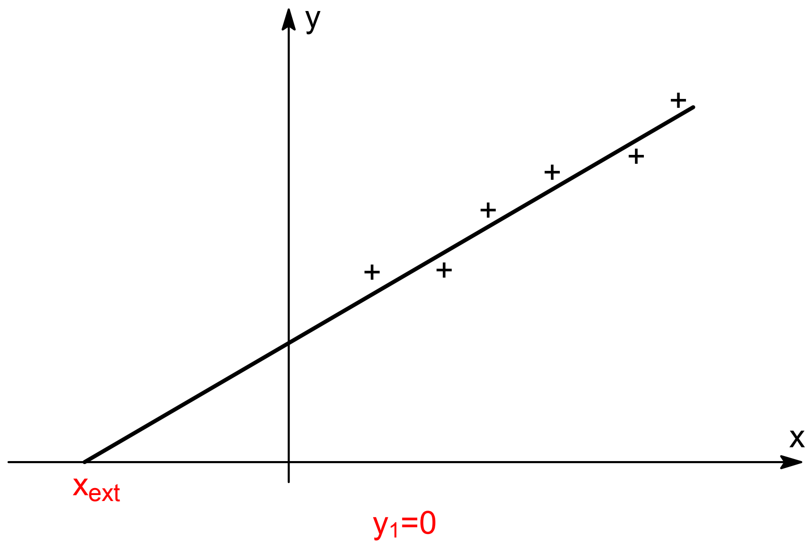
<u>Exercice RegLin4 :</u> Exploitation d'un dosage par étalonnage
Dosage des ions nitrite dans un jambon (x désigne la concentration initiale en ions nitrite et y l'absorbance de la solution obtenue).

	A	B	C	D	E	F	G
1							
2	**x**	**y**		**N**	**9,00**		
3	1	0,266		**Sx**	**47,00**		
4	2	0,327		**Sy**	**4,35**		
5	3	0,366		**Sx2**	**321,00**		
6	4	0,406		**Sy2**	**2,32**		
7	5	0,462		**Sxy**	**26,72**		
8	6	0,519		**D**	**680,00**		
9	7	0,542		**a0**	**2,06E-01**		
10	9	0,749		**a1**	**5,31E-02**		
11	10	0,712		**s(y)**	**0,03**		
12				**ymeas**	**0,460**		
13				**xcc**	**4,78**		
14							
15				**s(xcc)**	**0,61**		
16							
17			**xcc=**	**4,8**	**"+/-"**	**1,5**	
18							
19							
20							
21							
22							
23							
24							
25							
26							
27							
28							
29							
30							
31							
32							
33							
34							



N	=NBVAL(A3:A22)
Sx	=SOMME(A3:A22)
Sy	=SOMME(B3:B22)
Sx2	=SOMME.CARRES(A3:A22)
Sy2	=SOMME.CARRES(B3:B22)
Sxy	=SOMMEPROD(A3:A22;B3:B22)
D	=E2*E5-E3^2
a0	=(E4-E10*E3)/E2
a1	=(E2*E7-E3*E4)/E8
s(y)	=RACINE(1/(E2-2)*(E6-E10*E7-E9*E4))
ymeas	0,460
xcc	=1/E10*(E12-E9)
s(xcc)	=E11/E10*RACINE(1+1/E8*(E2*E13^2-2*E13*E3+E5))
Δxcc	=E15*LOI.STUDENT.INVERSE(0,05;E2-2)

- **Inter(extra)polation :** La fonction F est la fonction qui à une valeur particulière de $y(y_1)$ fait correspondre une valeur de x (x_{ext}) :



$$F = x_{ext} = \frac{y_1 - a_0}{a_1}$$

$$\frac{\partial x_{ext}}{\partial a_0} = -\frac{1}{a_1}$$

$$\frac{\partial x_{ext}}{\partial a_1} = -\frac{y_{1s} - a_0}{a_1^2} = -\frac{x_{ext}}{a_1}$$

$$\alpha_F^2 = \alpha_y^2 \frac{1}{D} \times \begin{vmatrix} -\frac{1}{a_1} & -\frac{x_{ext}}{a_1} \end{vmatrix} \times \begin{vmatrix} \sum_{i=1}^{N} x_i^2 & -\sum_{i=1}^{N} x_i \\ -\sum_{i=1}^{N} x_i & N \end{vmatrix} \times \begin{vmatrix} -\frac{1}{a_1} \\ -\frac{x_{ext}}{a_1} \end{vmatrix}$$

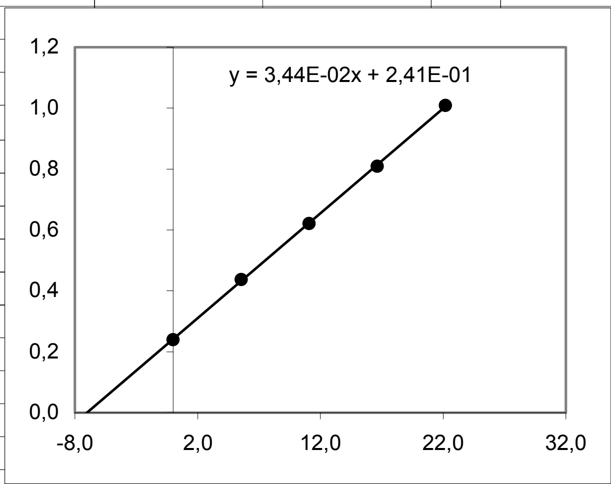
$$= \sigma_y^2 \frac{1}{D} \times \begin{vmatrix} -\frac{1}{a_1} \sum_{i=1}^{N} x_i^2 + \frac{x_{ext}}{a_1} \sum_{i=1}^{N} x_i & \frac{1}{a_1} \sum_{i=1}^{N} x_i - \frac{N x_{ext}}{a_1} \end{vmatrix} \times \begin{vmatrix} -\frac{1}{a_1} \\ -\frac{x_{ext}}{a_1} \end{vmatrix}$$

$$= \sigma_y^2 \frac{1}{D} \times \left[-\frac{1}{a_1} \sum_{i=1}^{N} x_i^2 + \frac{x_{ext}}{a_1} \sum_{i=1}^{N} x_i \right] \times \left(-\frac{1}{a_1} \right) - \frac{x_{ext}}{a_1} \times \left[\frac{1}{a_1} \sum_{i=1}^{N} x_i - \frac{N x_{ext}}{a_1} \right]$$

$$\boxed{\sigma_{x_{ext}}^2 = \sigma_F^2 = \sigma_y^2 \frac{1}{D} \times \frac{1}{a_1^2} \left[\sum_{i=1}^{N} x_i^2 - 2 x_{ext} \sum_{i=1}^{N} x_i + N x_{ext}^2 \right]}$$

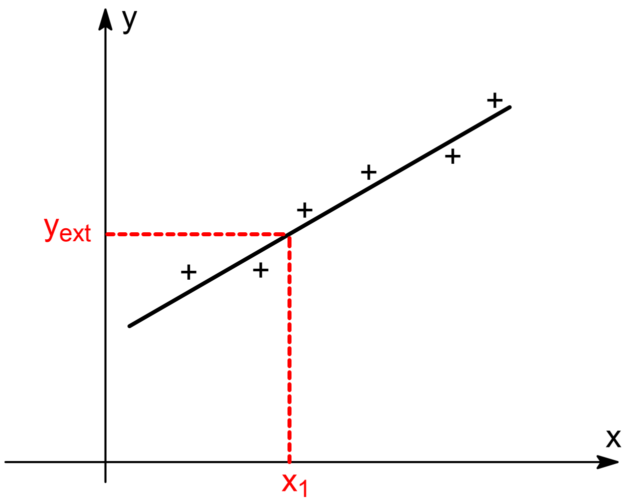
Exercice RegLin5 : Exploitation d'un dosage par la méthode des ajouts dosés

	A	B	C	D	E	F	G
1							
2	**x**	**y**		**N**	**5,00**		
3	0,00	0,240		**Sx**	**55,50**		
4	5,55	0,437		**Sy**	**3,12**		
5	11,10	0,621		**Sx2**	**924,08**		
6	16,65	0,809		**Sy2**	**2,31**		
7	22,20	1,009		**Sxy**	**45,19**		
8				**D**	**1540,13**		
9				**a0**	**2,41E-01**		
10				**a1**	**3,44E-02**		
11				**s(y)**	**4,86E-03**		
12				**xext**	**-7,009**		
13				**yext**	**0,000**		
14							
15				**s(yext)**	**1,59E-01**		
16							
17				**yext =**	**-7,01**	**"+/-"**	**0,51**
18							
19							
20							
21							
22							
23							
24							
25							
26							
27							
28							
29							
30							
31							
32							
33							



N	=NBVAL(A3:A22)
Sx	=SOMME(A3:A22)
Sy	=SOMME(B3:B22)
Sx2	=SOMME.CARRES(A3:A22)
Sy2	=SOMME.CARRES(B3:B22)
Sxy	=SOMMEPROD(A3:A22;B3:B22)
D	=E2*E5-E3^2
a0	=(E4-E10*E3)/E2
a1	=(E2*E7-E3*E4)/E8
s(y)	=RACINE(1/(E2-2)*(E6-E10*E7-E9*E4))
xext	=(E13-E9)/E10
yext	0
s(yext)	=RACINE(E11^2/E8/E10^2*(E5-2*E12*E3+E2*E12^2))
yext =	=E12 +/- E15*LOI.STUDENT.INVERSE(0,05;E2-2)

- **Inter(extra)polation :** La fonction F est la fonction qui à une valeur particulière de x (x_1) fait correspondre une valeur de y (x_1) :



$$F = y_{ext} = a_1 x_1 + a_0$$

$$\frac{\partial F}{\partial a_0} = 1$$

$$\frac{\partial F}{\partial a_1} = x_1$$

$$\sigma^2_{y ext} = \sigma^2_y \frac{1}{D} \times \begin{vmatrix} 1 & x_1 \end{vmatrix} \times \begin{vmatrix} \sum_{i=1}^{N} x_i^2 & -\sum_{i=1}^{N} x_i \\ -\sum_{i=1}^{N} x_i & N \end{vmatrix} \times \begin{vmatrix} 1 \\ x_1 \end{vmatrix}$$

$$= \sigma^2_y \frac{1}{D} \times \begin{vmatrix} \sum_{i=1}^{N} x_i^2 - x_1 \sum_{i=1}^{N} x_i & -\sum_{i=1}^{N} x_i + N x_1 \end{vmatrix} \times \begin{vmatrix} 1 \\ x_1 \end{vmatrix}$$

$$= \sigma^2_y \frac{1}{D} \times \left[\sum_{i=1}^{N} x_i^2 - x_1 \sum_{i=1}^{N} x_i \right] \times (1) + x_1 \times \left[-\sum_{i=1}^{N} x_i + N x_1 \right]$$

$$\boxed{\sigma^2_{y ext} = \sigma^2_F = \sigma^2_y \frac{1}{D} \times \left[\sum_{i=1}^{N} x_i^2 - 2 x_1 \sum_{i=1}^{N} x_i + N x_1^2 \right]}$$