

# S1 : Statistiques à deux variables.

Connaissances travaillées	Non acquis	En cours d'acquisition	Acquis
Représenter un nuage de points.			
Calculer le point moyen d'une série à deux variables.			
Déterminer un ajustement affine d'un nuage de points.			
Utiliser un ajustement affine pour prédire des valeurs.			
Utiliser un changement de variables.			

## I/ Nuage de points :

### 1) Qu'est ce qu'une série statistique à deux variables ?

Jusqu'ici vous aviez étudié des séries à une seule variable et vous avez appris à en calculer les différentes caractéristiques. Par exemple appris à calculer votre moyenne à partir de vos notes, si vous avez eu les résultats suivants : 11 ; 12 ; 13 ; 14 ; 15

Vous pouvez calculer que la moyenne est 13.

Dans ce chapitre les séries ne seront plus composées d'une seule mais de deux variables. Cela correspond à toutes les situations où on étudie l'évolution d'un caractère en fonction d'un autre. Par exemple lorsqu'en laboratoire vous comparez l'absorbance de plusieurs échantillons en fonction de leur concentration, vous obtenez un tableau contenant deux lignes, une avec l'absorbance des différentes fioles et l'autre avec leur concentration. Vous avez donc une série statistique à deux variables comme celle-ci :

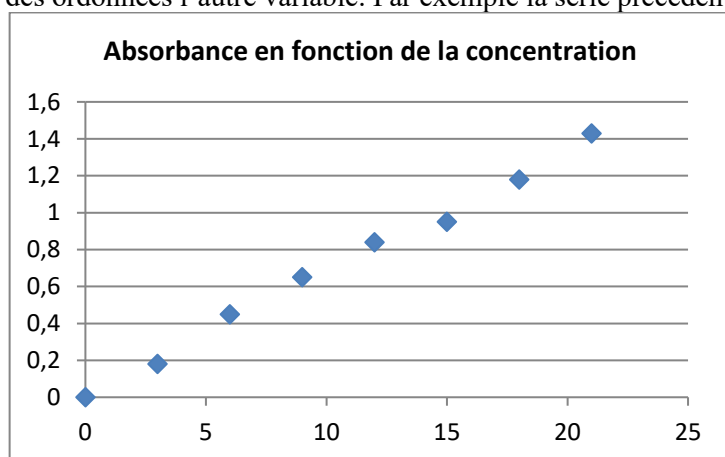
<b>Concentration (en mol.L<sup>-1</sup>)</b>	0	$3 \times 10^{-3}$	$6 \times 10^{-3}$	$9 \times 10^{-3}$	$12 \times 10^{-3}$	$15 \times 10^{-3}$	$18 \times 10^{-3}$	$21 \times 10^{-3}$
<b>Absorbance</b>	0	0,18	0,45	0,65	0,84	0,95	1,18	1,43

Généralement l'intérêt de ce genre de série est soit :

- De rechercher un lien entre deux variables, autrement dit trouver une formule permettant de calculer une valeur d'une variable à partir de l'autre.
- Tracer une droite d'étalonnage, on mesure l'évolution d'un facteur en fonction d'un autre que l'on maîtrise. Ce qui permet après de faire des extrapolations.

### 2) Représenter un nuage de points :

Pour représenter un nuage de point on peut utiliser un repère et placer sur l'axe des abscisses une des variables et sur l'axe des ordonnées l'autre variable. Par exemple la série précédente donne ce nuage de points.



Pour placer les points il suffit de prendre colonne par colonne, la première ligne correspond à l'abscisse du point et la seconde à l'ordonnée. Ainsi on peut placer le point de coordonnée (3 ; 0,18)

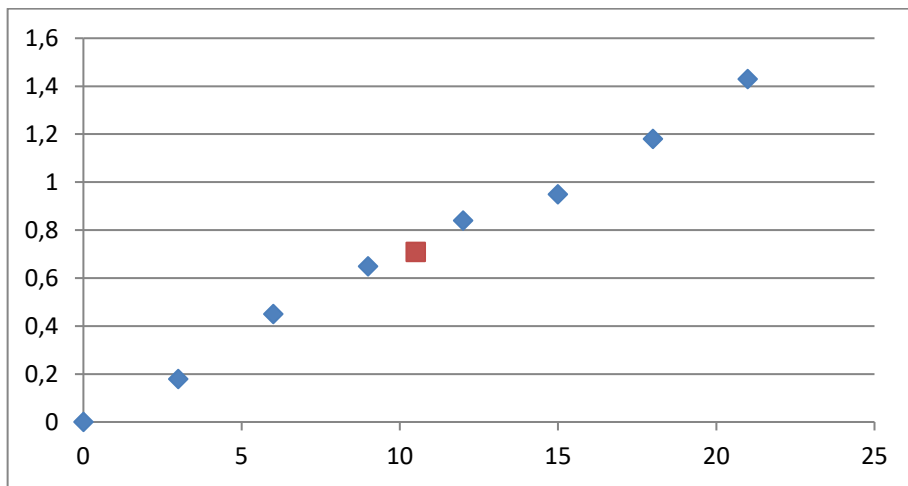
## II/ Point moyen et ajustement affine :

### 1) Point moyen :

Le point moyen est un point qui a la même utilité qu'une moyenne il sert à résumer une série de deux variables à l'aide de deux valeurs. Pour en calculer les coordonnées il suffit de faire la moyenne des valeurs en abscisse et la moyenne des valeurs en ordonnées.

Par exemple pour la série précédente, le point moyen a pour coordonnées :

$$x_m = \frac{0 + 3 + 6 + 9 + 12 + 15 + 18 + 21}{8} = 10,5 \text{ et } y_m = \frac{0 + 0,18 + 0,45 + 0,65 + 0,84 + 0,95 + 1,18 + 1,43}{8} = 0,71$$

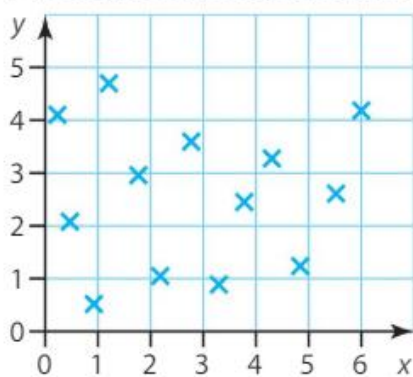


Le point moyen est en rouge sur le graphique, on remarque qu'il est « au milieu » du nuage de points.

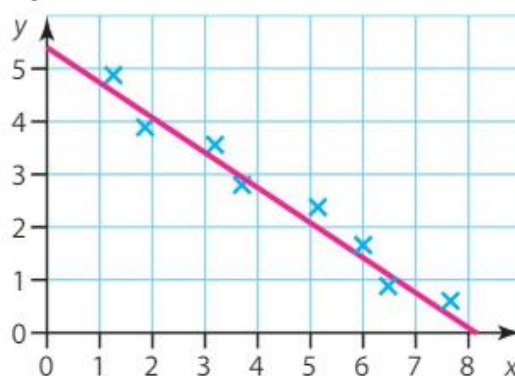
## 2) Ajustement affine :

Les nuages de points peuvent permettre de déterminer s'il existe une relation particulière entre les deux variables étudiées. Selon l'apparence du nuage on peut envisager plusieurs types de relations.

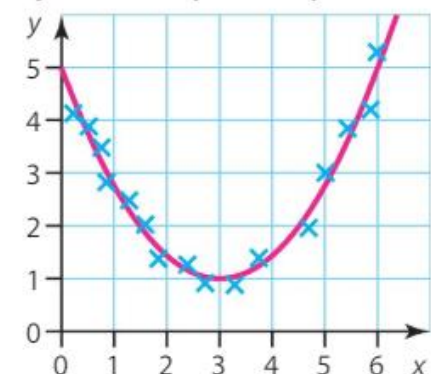
### Pas d'ajustement particulier



### Ajustement affine



### Ajustement par une parabole



Dans le premier cas, le nuage de point ne semble prendre aucune forme particulière, on peut imaginer qu'aucune relation ne lie donc les deux variables. Dans le second cas les points semblent presque alignés, on peut donc imaginer que la relation qui lie les deux variables est une fonction affine. Et dans le troisième cas, les points semblent dessiner une parabole, on peut donc imaginer que les deux variables sont reliées par une fonction de degré 2. Cette année nous n'aborderons que des ajustements affines.

Pour réaliser un tel ajustement il faut utiliser la méthode des moindres carrés, c'est un procédé de calcul qui vise à chercher l'équation de la droite telle de manière à ce que la distance entre la droite et les points du nuage soit le plus faible possible. Ce procédé de calcul pouvant être complexe nous ferons appel à la calculatrice.

### Méthode :

Le nuage de point tracé donnant l'absorbance en fonction de la concentration ressemblait à une droite, nous allons donc utiliser la calculatrice pour déterminer l'équation de la droite d'ajustement affine de cette série

**Etape 1 :** Entrer les valeurs de la série dans la calculatrice :

Presser la touche :




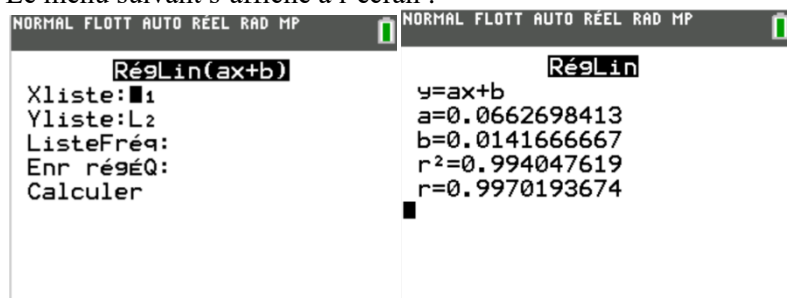
Puis sélectionner modifier, dans les colonnes L1 et L2 entrer vos deux séries de valeur.

L1	L2	L3	L4	L5	2
0	0	---	---	---	
3	0.18				
6	0.45				
9	0.65				
12	0.84				
15	0.95				
18	1.18				
21	1.43				
---	---				

L2(9)=

## Etape 2 : Calculer la droite d'ajustement :

Presser la touche  puis sélectionner le menu calc. Dans ce menu choisir la 4ème fonction Reglin(ax+b). Le menu suivant s'affiche à l'écran :



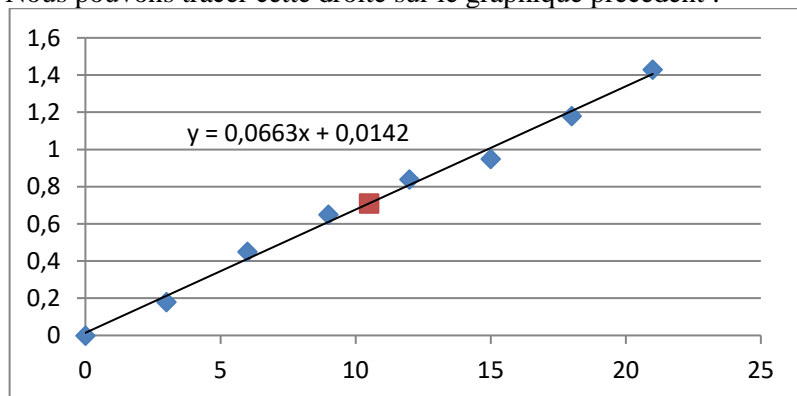
Le paramètre XListe doit être le nom de la liste que vous avez utilisé pour les abscisses sur votre nuage de points et Yliste celle que vous avez utilisé pour les ordonnées. Les autres paramètres doivent rester vides. Une fois que vous avez sélectionné Calculer et vous obtenez le résultat de l'image de droite.

On peut y lire les valeurs de a et de b qui nous permettront de reconstituer l'équation de droite. Ainsi qu'un paramètre nommé  $r^2$ , pour simplifier ce paramètre indique à quel point l'ajustement choisit modélise bien la série. Plus il est proche de 1, plus notre ajustement est bon.

Grâce à la calculatrice nous en déduisons que la droite d'ajustement affine de notre série de donnée à pour équation :

$$y = 0,0663x + 0,0142$$

Nous pouvons tracer cette droite sur le graphique précédent :



Nous remarquons que la droite d'ajustement passe par le point moyen. Nous admettrons que :

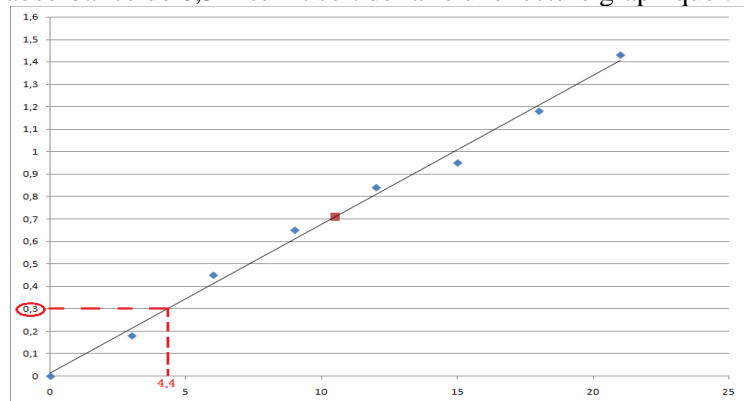
### **La droite d'ajustement passe toujours par le point moyen du nuage.**

La constatation précédente peut nous permettre de tracer au jugé la droite d'ajustement d'un nuage, en déterminant le point moyen puis en traçant la droite qui nous semble passer au plus proche de tous les points du nuage. Nous noterons cependant que cette méthode est bien moins précise que la calculatrice.

## III/ Utilisation de l'ajustement affine :

Comme nous l'avons dit précédemment l'intérêt de l'ajustement (qu'il soit affine ou d'une autre sorte) est d'obtenir une expression (une formule) permettant de calculer la valeur d'une variable en connaissant une valeur précise de l'autre variable.

Si on reprend l'exemple précédent et que l'on cherche à calculer par exemple la concentration correspondant à une absorbance de 0,3 il suffit soit de faire une lecture graphique :



Soit d'utiliser l'équation de droite précédemment calculée :

$$y = 0,0663x + 0,0142$$

On sait que l'absorbance (donc  $y$ ) est de 0,3 et on cherche la concentration (donc  $x$ ) :

$$0,3 = 0,0663x + 0,0142$$

$$0,3 - 0,0142 = 0,0663x + 0,0142 - 0,0142$$

$$0,2858 = 0,0663x$$

$$\frac{0,2858}{0,0663} = \frac{0,0663x}{0,0663}$$

$$4,311 \approx x$$

On remarque que les résultats ont assez proches l'un de l'autre. Bien entendu celui obtenu par le calcul est bien plus précis cependant selon la situation la détermination graphique peut s'avérer suffisante.

#### IV/ Ajustement affine et changement de variables :

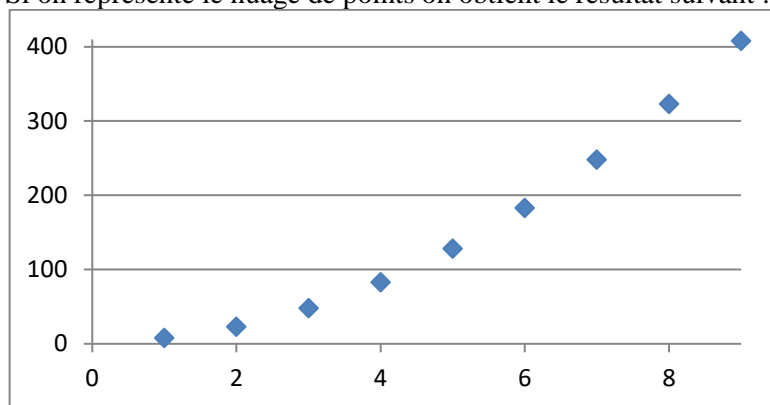
Dans certains cas le nuage de point ne semble pas former une droite, mais une forme générale semble quand même se dessiner. Il est alors possible d'effectuer ce que l'on appelle un changement de variable afin d'obtenir une droite à partir de laquelle on fait un ajustement affine.

##### Exemple :

Soit la série suivante (chiffres inventés et peu réalistes) :

Age	1	2	3	4	5	6	7	8	9
Nombre de mots connus	8	23	48	83	128	183	248	323	408

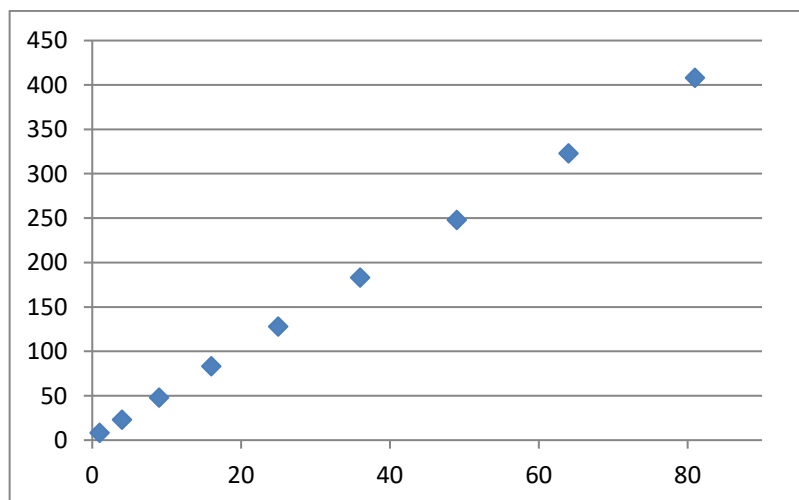
Si on représente le nuage de points on obtient le résultat suivant :



On remarque que les points ne sont pas alignés. Mais qu'une tendance générale se dégage. Il semble bien qu'il y ait une relation entre les deux variables, mais cette relation ne semble pas être affine.

Nous allons changer la variable ici, pour cela nous allons remplacer  $x$  par  $x^2$  (le choix du calcul vous sera toujours donné lors d'un changement de variables) puis tracer  $y$  en fonction de  $x^2$  :

Age	1	2	3	4	5	6	7	8	9
Age <sup>2</sup>	1	4	9	16	25	36	49	64	81
Nombre de mots connus	8	23	48	83	128	183	248	323	408



Le nouveau nuage de point est bien une droite dont on peut calculer l'ajustement affine. Ici on obtient  $y = 5x + 3$ . Cela signifie que le nombre de mots connus peut être calculé à partir du carré de l'âge. On en déduit donc que :

$$\text{Nombre de mots connus} = 5 \times \text{âge}^2 + 3$$

Pour résumer un changement de variable c'est tout simplement effectuer un calcul avec les données en abscisse (ou en ordonnée), ensuite on recalcule l'ajustement à partir de cette nouvelle variable.