

Chapitre 2. Statistiques à 2 variables

Yann Barsamian

École Européenne de Bruxelles 1

Année scolaire 2023–2024



- Les nuages de points
- Les ajustements
- Interpolation et extrapolation

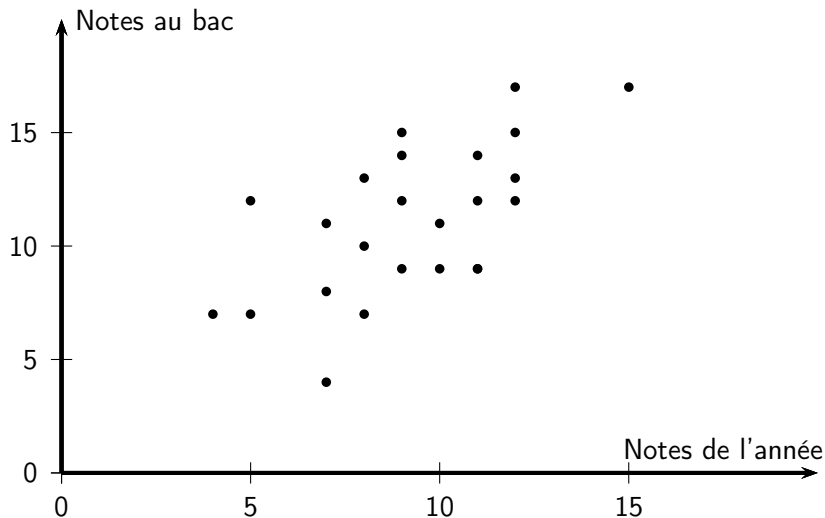
Dans une série statistique à deux variables, on s'intéresse à deux caractères x et y .

Exemple 1 : pour une classe de Terminale en mathématiques, on regarde leur moyenne x au long de l'année et leur note y au baccalauréat (sur 20). Voici des notes fictives pour une classe de 24 élèves. Ainsi, chaque couple (x_i, y_i) (pour i de 1 à 24) représente les notes d'un élève de la classe.

x_i	8	9	7	15	12	12	10	8	11	11	7	8
y_i	7	9	4	17	13	15	9	13	14	9	11	10
x_i	11	11	12	12	7	9	9	5	9	5	10	4
y_i	9	12	17	12	8	15	12	7	14	12	11	7

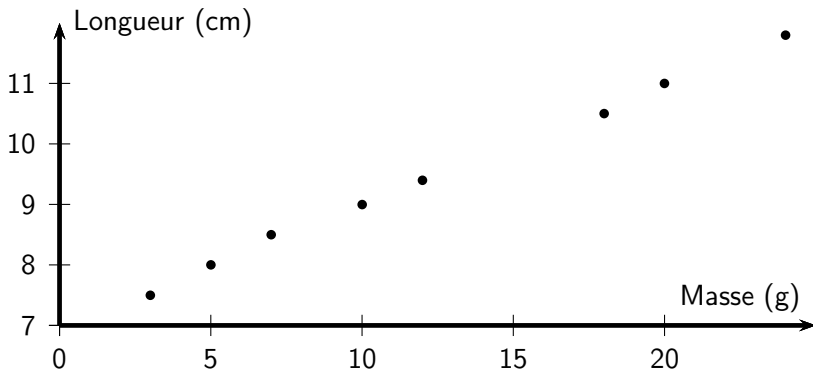
I/ Nuages de points

Pour visualiser les notes de ces élèves, on réalise un nuage de points.
Voir <https://www.youtube.com/watch?v=Nn6uckb3RvE>.



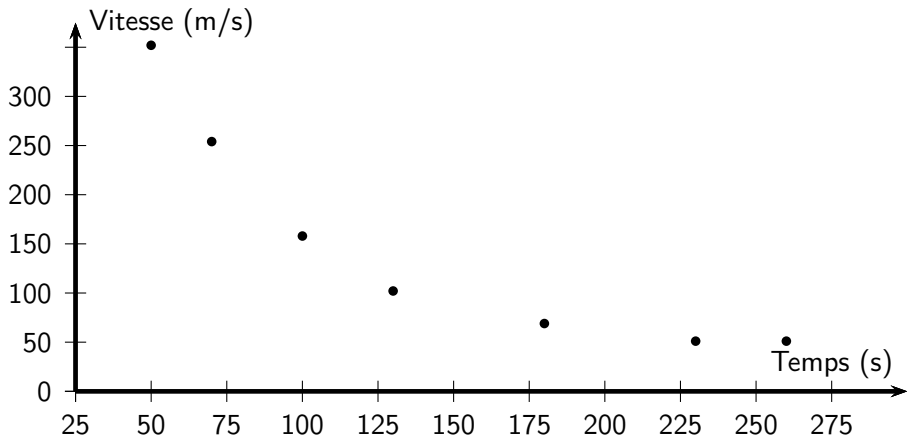
Second exemple : on attache différents objets à un ressort, et on regarde, pour chaque objet, la masse x de l'objet attaché, et la longueur y du ressort quand cet objet y est attaché.

x_i	7	10	18	20	5	24	12	3
y_i	8.5	9	10.5	11	8	11.8	9.4	7.5

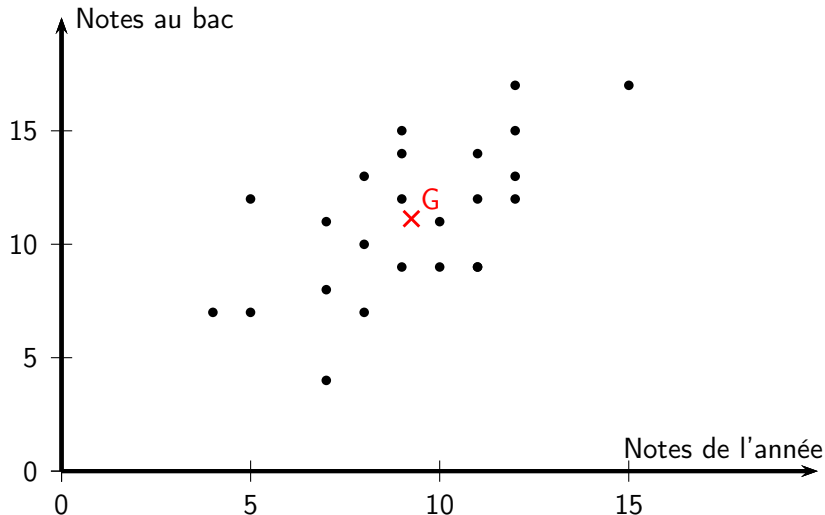


Troisième exemple : on a mesuré, pour un sauteur en parachute, à différents temps x depuis le saut, sa vitesse y .

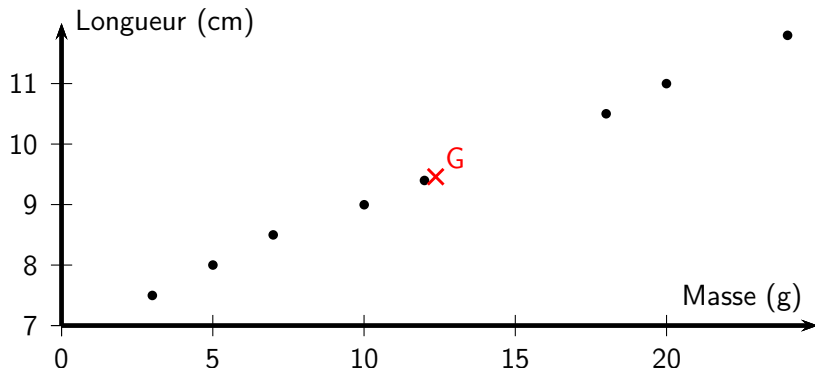
x_i	50	70	100	130	180	230	260
y_i	352	254	158	102	69	51	51



On se servira beaucoup, dans ce chapitre, du point moyen du nuage. C'est tout simplement le point qui a pour coordonnées (\bar{x}, \bar{y}) .

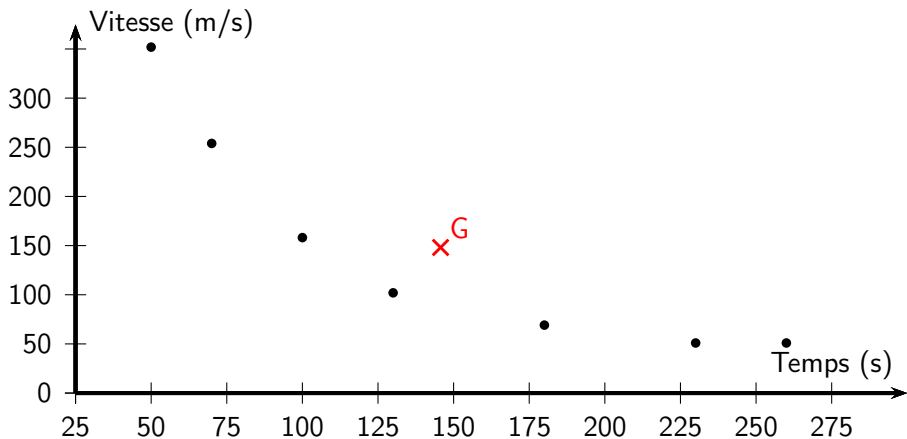


On se servira beaucoup, dans ce chapitre, du point moyen du nuage. C'est tout simplement le point qui a pour coordonnées (\bar{x}, \bar{y}) .



I/ Nuages de points

On se servira beaucoup, dans ce chapitre, du point moyen du nuage. C'est tout simplement le point qui a pour coordonnées (\bar{x}, \bar{y}) .



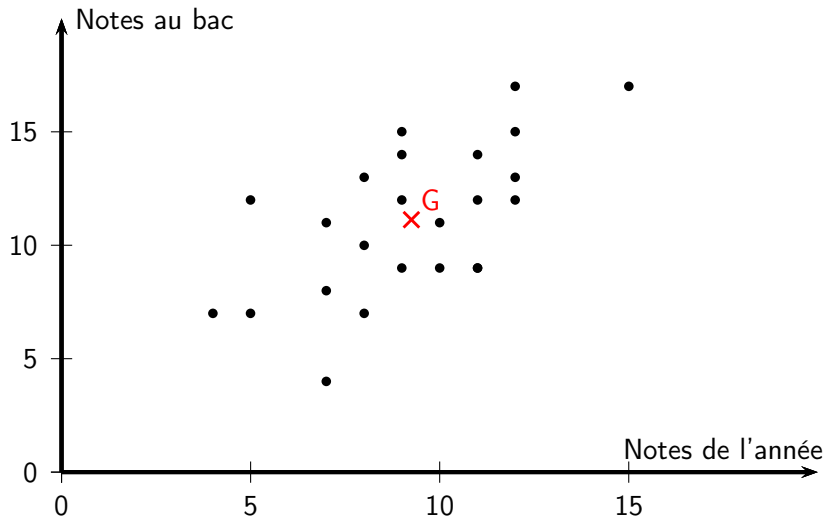
Il est utile d'essayer d'exprimer une variable en fonction d'une autre. On appellera variable dépendante la variable "explicative", et l'autre sera la variable indépendante. On essaye donc d'exprimer la variable indépendante en fonction de l'autre. Typiquement, on essaiera d'exprimer y en fonction de x .

Pour cela, on va donc essayer de tracer une fonction $f(x)$ qui soit ressemblante au nuage de points. On va démarrer par les fonctions les plus simples : les fonctions affines (des droites).

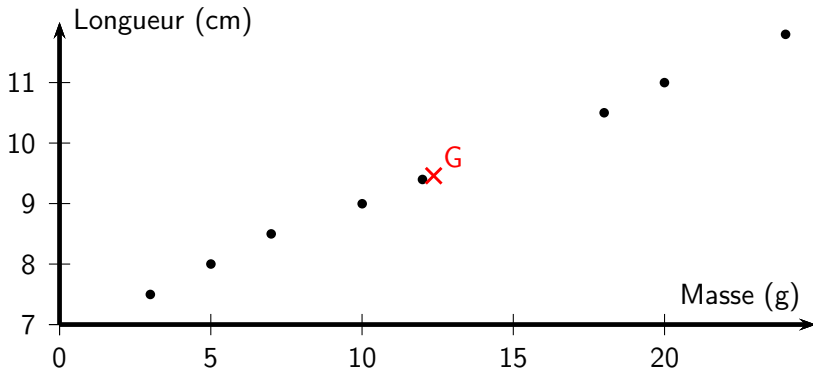
À partir de maintenant, tout est à la calculatrice ! Pour le prébac, vous n'avez rien à savoir faire à la main. Pour le baccalauréat, on verra d'autres méthodes dont l'une doit savoir être faite à la main.

Par contre, il faut toujours garder son sens critique : est-ce que la fonction que je suis en train de tracer a l'air de bien correspondre au nuage de points ?

Traçons à la main des ajustements affines des 3 nuages de points donnés en exemple.

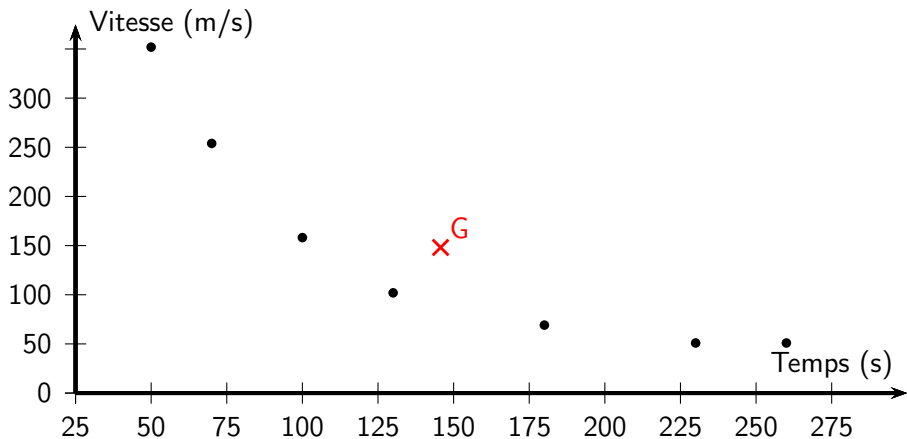


Traçons à la main des ajustements affines des 3 nuages de points donnés en exemple.

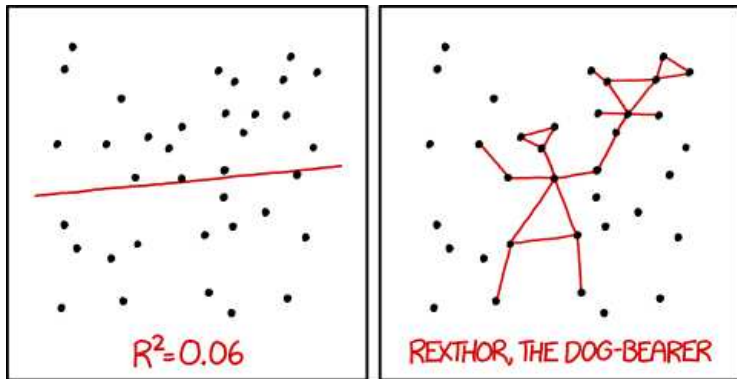


II/ Ajustements

Traçons à la main des ajustements affines des 3 nuages de points donnés en exemple.



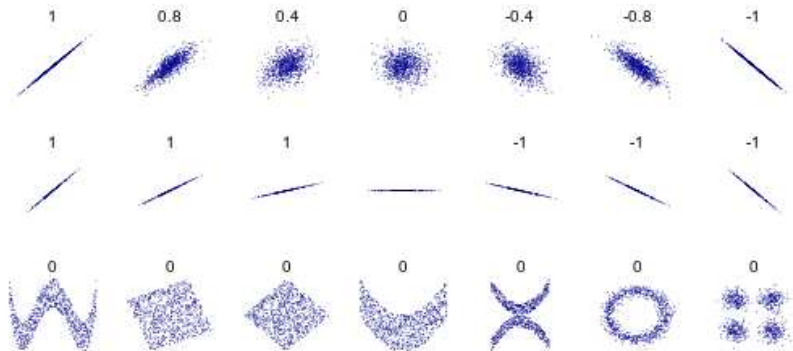
Le coefficient de corrélation linéaire de Bravais–Pearson, noté r .



I DON'T TRUST LINEAR REGRESSIONS WHEN IT'S HARDER
TO GUESS THE DIRECTION OF THE CORRELATION FROM THE
SCATTER PLOT THAN TO FIND NEW CONSTELLATIONS ON IT.

Source : <https://xkcd.com/1725/>

Le coefficient de corrélation linéaire de Bravais–Pearson, noté r .



Source : [https://fr.wikipedia.org/wiki/Corr%C3%A9lation_\(statistiques\)](https://fr.wikipedia.org/wiki/Corr%C3%A9lation_(statistiques)).

Au niveau de la calculatrice, il faut savoir faire les points 1 à 4 de la vidéo de tutoriel suivante :

<https://www.youtube.com/watch?v=Nb2hmeoaMLo> (le lien se trouve sur la page de notre cours, tout en haut dans le programme officiel : <http://www.barsamian.am/2023-2024/S7P3/>).

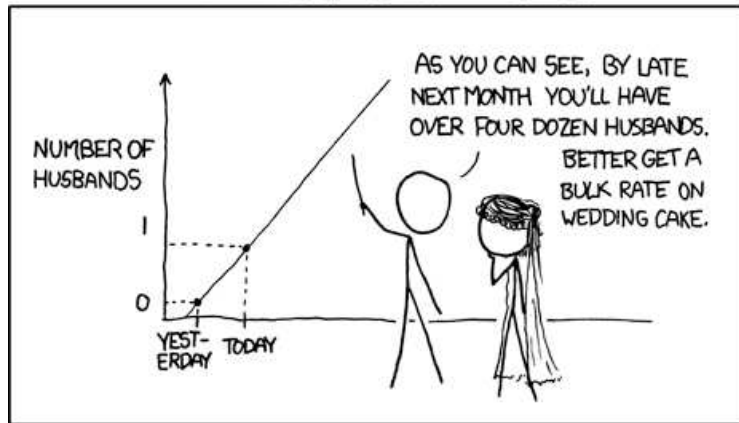
- Rentrer une colonne pour x_i , une colonne pour y_i .
- L'outil "Statistiques à 2 variables" donne le point moyen et r .
- L'outil "Régression linéaire" calcule l'équation de la droite de régression par la méthode des moindres carrés $y = mx + p$ (ainsi que r).
- L'outil "Données et statistiques" permet d'afficher le nuage de points, conjointement avec la droite de régression.

Une fois qu'on a fait un modèle à partir des données (ici, le modèle affine, via une droite de régression), on peut utiliser le modèle pour :

- Interpoler : trouver une valeur à l'intérieur du nuage (mais pas déjà connue par un point).
- Extrapoler : trouver une valeur à l'extérieur du nuage.

Attention quand on extrapole...

MY HOBBY: EXTRAPOLATING



Source : <https://xkcd.com/605/>.