TSTMG ∼ Séries statistiques à deux variables

1 - Introduction

Voici deux séries statistiques représentant deux situations distinctes. La première donne l'évolution du chiffre d'affaire d'une entreprise au cours du temps (l'année de rang 0 étant 2010) et la deuxième la distance de freinage d'un véhicule en fonction de la vitesse.

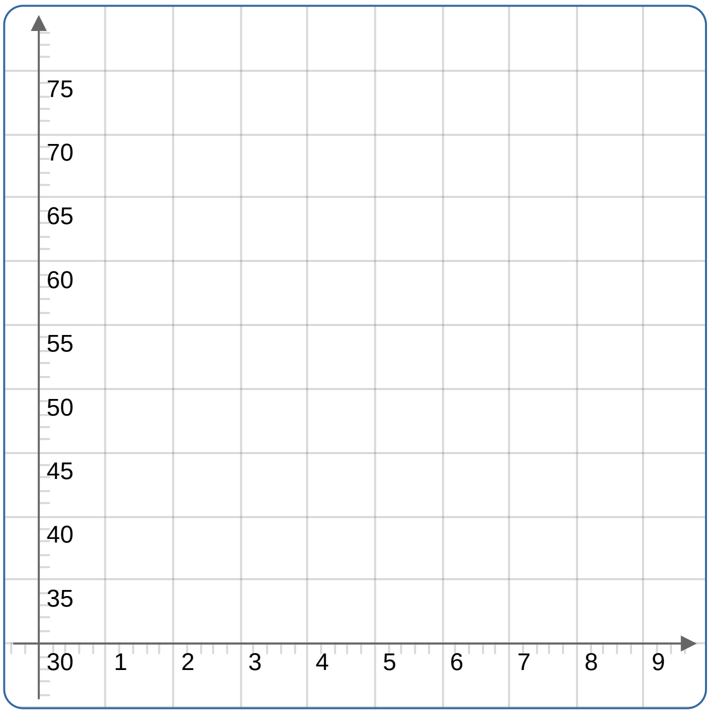
Rang de l'année	0	1	2	3	4	5	6	7	8
CA en milliers d'euros	$45,3$	$42,2$	$49,8$	$52,1$	$58,8$	$\ldots$	$60,5$	$64,1$	$66,4$

Évolution du chiffre d'affaire d'une entreprise

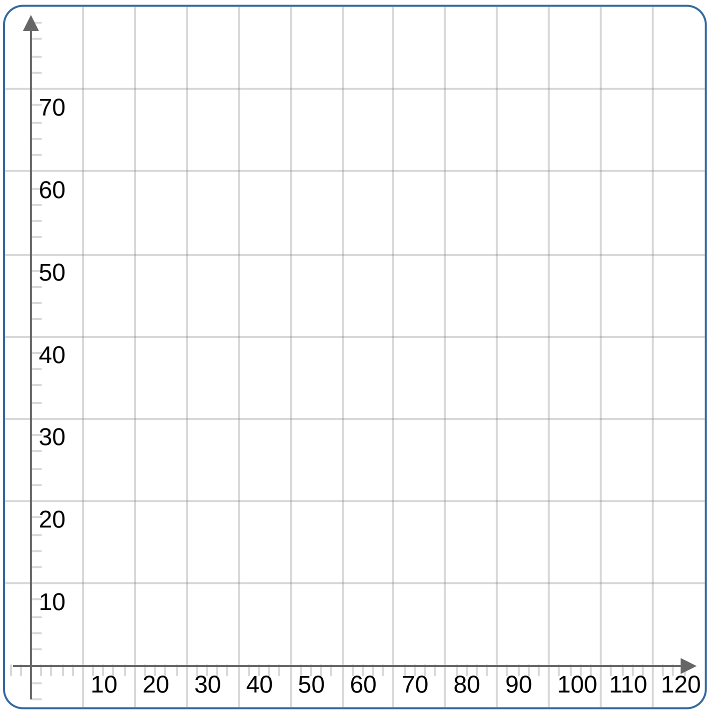
Vitesse en km/h	20	30	40	50	60	70	80	90	100	110	120
Distance de freinage en m	2	5	9	12	$\ldots$	24	32	41	50	$\ldots$	72

Distance de freinage d'un véhicule

Construire dans les repères ci-dessous les nuages de points associés à ces deux séries, puis donner une estimation des valeurs manquantes dans ces deux tableaux.



Évolution du chiffre d'affaire d'une entreprise



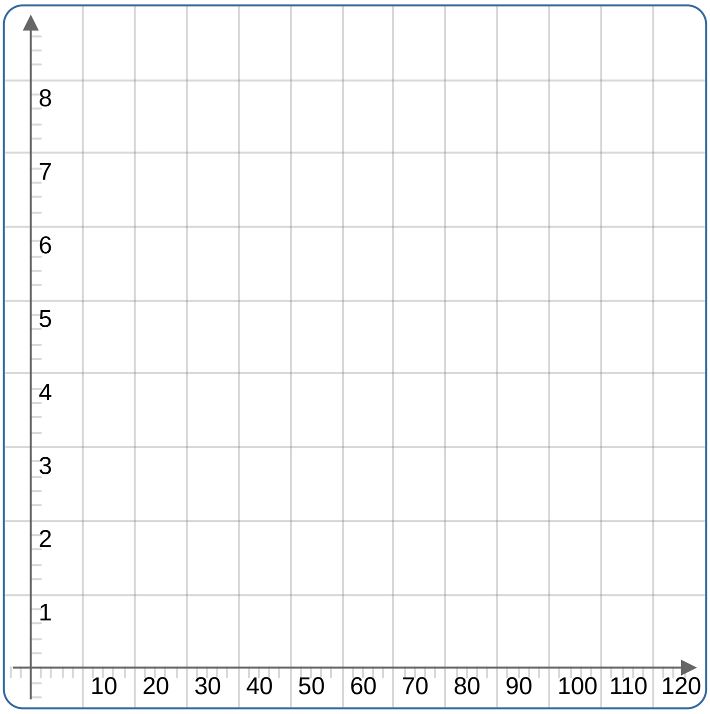
Distance de freinage d'un véhicule

Remarque 1

Lorsque le nuage de points a une forme   on peut l'approcher par une   , et ainsi avoir une idée des valeurs manquantes ou de futures valeurs. Les droites sont des objets dont on sait trouver assez facilement des équations.
Dans la deuxième situation, le nuage de points   et on ne peut pas trouver une   qui passent   des points qui le compose. Il faudrait trouver dans ce cas précis une   dont l'équation n'est pas aisée à déterminer. On se propose dans cette situation de modifier les données pour obtenir un nouveau nuage de points cette fois-ci de forme

Vitesse en km/h	20	30	40	50	60	70	80	90	100	110	120
Racine carrée de la distance	$1,41$	$2,24$	3	$3,46$	$\ldots$	$4,90$	$5,66$	$6,40$	$7,07$	$\ldots$	$8,49$

Construisons le graphique associé :



Sur ce graphique, on peut tracer une droite qui passe au plus proche de ces nouveaux points. On trouve par exemple que pour une vitesse de 60 km/h, la racine carrée de la distance peut être approchée par . Il suffit dans ce cas d'élever au carré pour retrouver la distance de freinage correspondante, soit

2 - Séries statistiques à deux variables

Sur une même population, on peut étudier plusieurs caractères quantitatifs : le chiffre d'affaire d'une entreprise en fonction des années, ou la distance de freinage d'un véhicule en fonction de la vitesse initiale, ou encore le nombre de bactéries dans une solution en fonction de la température du milieu, ou bien la charge de rupture de tiges en acier en fonction de leur teneur en carbone etc.
Le but est de déterminer si il existe un entre les deux caractères étudiés.

Définition 1 -- *Séries statistiques à deux variables*
Soient x et y deux caractères d'une population.
À chaque individu de la population, on associe un couple où x_i et y_i sont des valeurs prises par les caractères x et y.
Une **série statistique à deux variables** est

Exemple 1

Les deux tableaux donnés en introduction représentent deux séries statistiques.

Définition 2 -- *Nuage de points*
Soit une série statistique à deux variables x et y prenant respectivement les valeurs $x_1, x_2, \ldots, x_n$ et $y_1, y_2, \ldots, y_n$. Le plan étant muni d'un repère, on associe au couple
Le associé à la série statistique est l'ensemble des ainsi obtenus.

Exemple 2

Les graphiques construits en introduction sont des nuages de points associés à chacune des séries statistiques étudiées.

Définition 3 -- *Point moyen*
Soit une série statistique à deux variables x et y prenant respectivement les valeurs x_1, x_2, . . ., x_n et y_1, y_2, . . ., y_n.
Le du nuage statistique est le point

Exemple 3

Dans la série statistique associée au chiffre d'affaire de l'entreprise au cours du temps, le point moyen est :

3 - Ajustement affine

Définition 4 -- *Ajustement affine*
Étant donné une série statistique double et son nuage de points, on peut chercher une fonction f dont la courbe représentative $\mathcal{C}$ passe des points du nuage.
Le problème de l'ajustement consiste à déterminer cette fonction f.
L'ajustement est dit lorsque le graphe $\mathcal{C}$ de cette fonction est

Définition 5 -- *Ajustement au jugé*
Si les points du nuage statistique d'une série double semblent alignés, on peut tracer c'est-à-dire, visuellement au plus proche des points du nuage.

Exemple 4

C'est la méthode que nous avons employée en introduction.

Remarque 2

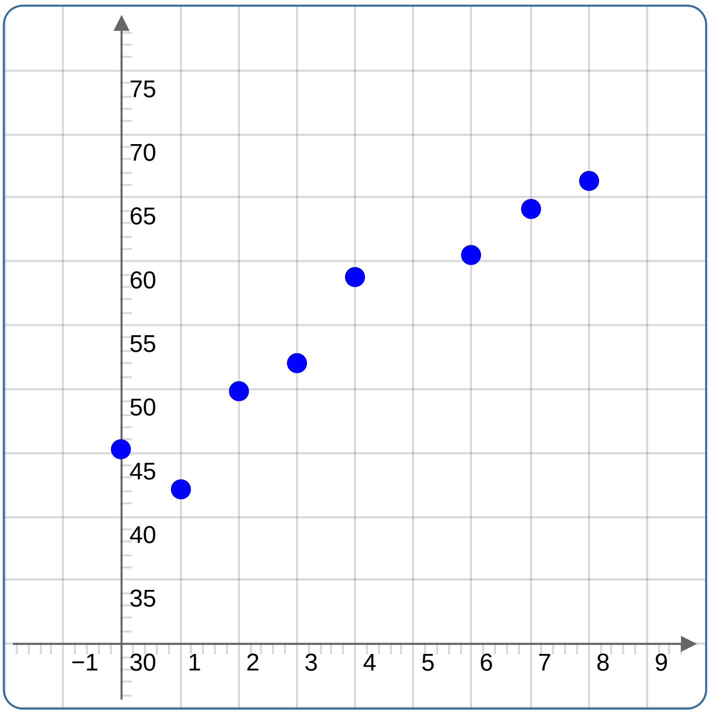
Lorsqu'on effectue un ajustement au jugé il y a une de possiblités et le choix de la droite d'ajustement n'est donc pas

Définition 6 -- *Méthode des moindres carrés*
On considère une série statistique double et son nuage de points $M_i(x_i\,;y_i)$ avec i compris entre 1 et n.
Soit une droite d d'équation À chaque point du nuage de points, on associe le point de la droite d.
Pour chaque i, on calcule et on les ajoute pour obtenir :

La méthode des consiste à déterminer les valeurs de pour que la somme E soit

Exemple 5

En reprenant la série statistique représentant l'évolution du chiffre d'affaire, on a le graphique suivant :



Évolution du chiffre d'affaire d'une entreprise

Propriété 1
Soit une série statistique à deux variables x et y prenant respectivement les valeurs x_1, x_2, $\ldots$, x_n et y_1, y_2, $\ldots$, y_n.
La droite obtenue par la méthode des moindres a pour équation $y = ax + b$ avec :

Remarque 3

Cette formule n'est pas à apprendre, les valeurs de a et b seront obtenues en manipulant la calculatrice. On peut d'ailleurs l'obtenir également à l'aide d'un tableur.

Propriété 2
La droite des moindres carrés passe par du nuage de points.

Exemple 6

Dans l'exemple de l'évolution du chiffre d'affaire, on obtient pour équation de la droite d'ajustement affine (en appliquant la méthode des moindres carrées à la calculatrice) :
Ainsi, pour l'année de rang 5 (soit en), avec cet ajustement on obtient un chiffre d'affaire de :