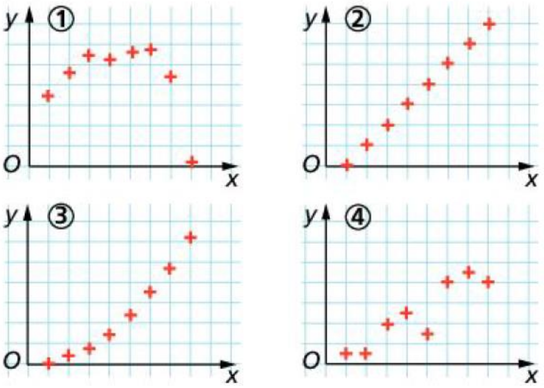
Statistiques à deux variables

Nuage de points et ajustement affine

10 ★ On considère les cinq séries du tableau suivant et quatre nuages de points :

x_i	1	2	3	4	5	6	7	8
y_i	1	1	4	5	3	8	9	8
z_i	0	3	8	15	24	35	48	63
t_i	0	10	20	30	40	50	60	70
w_i	6,71	9,29	10,71	10,2	11,6	11,96	8,64	0,23



1. Associer à chaque nuage la série statistique double qu'il représente parmi les 4 séries suivantes :

$(x_i ; y_i)$, $(x_i ; z_i)$, $(x_i ; t_i)$ ou $(x_i ; w_i)$.

On pourra visualiser ces séries doubles sur la calculatrice.

2. Indiquer les nuages de points pour lesquels un ajustement affine est judicieux. On pourra calculer les coefficients de corrélation linéaire.

11 ★ Depuis les années 1980, la consommation de médicaments en France a explosé.

La dépense annuelle, en euros par personne, tous les 5 ans, est donnée dans le tableau ci-après :

Année x_i	1980	1985	1990	1995	2000	2005	2010
Dépense y_i	95	177	258	331	414	500	525

1. À la calculatrice, ou sur papier, représenter le nuage de points $M_i(x_i ; y_i)$: un ajustement affine est-il justifié ? Déterminer le point moyen.

2. a) Déterminer l'équation de la droite d'ajustement $\mathscr{D}$ de y en x par la méthode des moindres carrés, sous la forme $y = ax + b$.

On donnera les coefficients a et b arrondis à l'unité.

b) Indiquer une interprétation du coefficient b.

3. On réalise le changement de variable $x' = x - 1980$, donnant le rang de l'année depuis 1980.

Pour cela, on peut utiliser l'instruction :

L1 – 1980 sto→ L3

a) Déterminer l'équation de la droite de régression $\mathscr{D}'$ de y en x' sous la forme $y = m x' + p$.

Comparer a et m. Interpréter ce coefficient.

b) Interpréter le coefficient p.

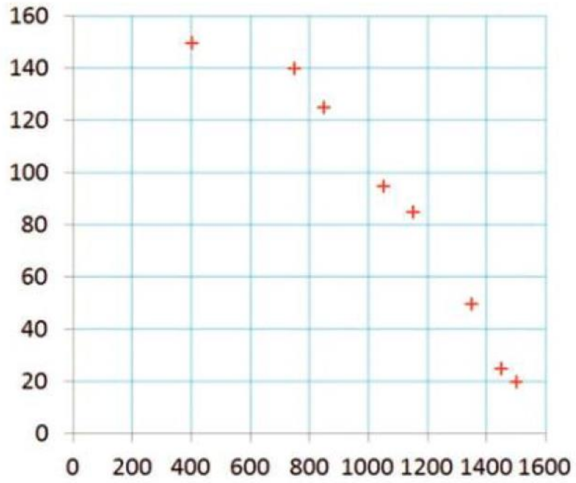
c) Extrapoler la dépense en médicaments par personne en 2020 suivant cet ajustement.

12 ★ Avant de proposer une croisière, une agence de voyages étudie le nombre de clients qui seraient prêts à réserver ce nouveau produit.

Elle obtient les résultats ci-dessous :

Prix en euros x_i	400	750	850	1050	1150	1350	1450	1500
Nombre de clients y_i	150	140	125	95	85	50	25	20

1. On a représenté le nuage de points sur tableur :



a) Reproduire le graphique sur papier millimétré.

b) À la calculatrice, calculer les coordonnées du point moyen $G(\bar{x} ; \bar{y})$. Placer ce point sur le graphique.

2. On choisit de réaliser un ajustement du nuage par la droite $\mathscr{D}_1$, dite « **droite des extrêmes** » passant par les points $M_1(400 ; 150)$ et $M_8(1\,500 ; 20)$.

a) Déterminer l'équation réduite de la droite $\mathscr{D}_1$.

Arrondir le coefficient a au millième.

b) Le point moyen appartient-il à cette droite ?

c) Tracer cette droite sur le graphique précédent.

Cette droite fournit-elle un bon ajustement du nuage de points ? Argumenter.

3. Soit $\mathscr{D}$ la droite d'ajustement de cette série statistique double, par la méthode des moindres carrés.

À la calculatrice, déterminer l'équation de $\mathscr{D}$, puis tracer cette droite sur le graphique.

Vérifier que cette droite passe par le point moyen.

4. En utilisant l'ajustement par la droite $\mathscr{D}$, calculer par interpolation le prix à partir duquel moins de 100 personnes seraient prêtes à réserver cette croisière.

13 ★ On considère la série statistique ci-dessous : x est le prix des framboises, en euros par kg , et y la quantité vendue correspondante, en centaines de kg.

Prix x_i	5	5,7	7,1	8,6	9	9,5	12
Quantité y_i	12	11	9	7,2	6,5	6	2

1. a) Sur la calculatrice, entrer en **liste 1** la série des prix et en **liste 2** celle des quantités.

b) Représenter le nuage de points sur l'écran de la calculatrice.

c) Calculer le prix moyen, arrondi au centime d'euro, et la quantité moyenne, arrondie à 10 kg .

2. Calculer le chiffre d'affaires engendré par la vente de $\overline{y}$ centaines de kg de framboises au prix moyen $\overline{x}$.

3. a) Un ajustement affine du nuage de points est-il judicieux ? Justifier en calculant le coefficient de corrélation linéaire.

b) Déterminer l'équation de la droite de régression de y en x . Arrondir les coefficients au dixième.

c) En utilisant ce modèle, calculer une estimation du chiffre d'affaires engendré par la vente d'une tonne de framboises.

14 ★ On considère la série statistique à deux variables suivante, saisie sur tableur :

	A	B	C	D	E	F	G	H	I	J
1	x_i	8	15	11	16	23	19	23	5	
2	y_i	25	30	30	42	50	45	44	18	

1. Sur une feuille de tableur, visualiser le nuage de points associé à cette série statistique double.

2. a) Calculer en **cellule J1** la moyenne $\overline{x}$ des valeurs x_i et en **cellule J2** celle des valeurs y_i .

b) Sélectionner les données situées en **J1** et **J2** et les ajouter sur le graphique précédent afin de représenter le point moyen sur le graphique.

3. a) En **cellule A3**, calculer le coefficient de corrélation linéaire de cette série double. Un ajustement affine du nuage de points est-il judicieux ?

b) Afficher sur le graphique la droite de régression de y en x, ainsi que son équation.

15 ★★ On désire étudier **l'influence d'un changement de variable.**

Le tableau suivant indique le nombre, en milliers, de nouveaux retraités, tous régimes confondus, au 31 décembre de l'année :

	A	B	C	D	E	F	G	H
1	Année, a_i	2004	2005	2006	2007	2008	2009	2010
2	Rang , x_i	0	1	2	3	4	5	6
3	Nouveaux retraités , y_i	696	673	763	796	811	704	736
4	zi							

1. Sur une feuille de tableur, représenter le nuage de points associé à la série statistique double (a_i ; y_i) .

2. Le rang de l'année 2004 est égal à 0 .

Indiquer la formule à saisir en **cellule B2**, et à recopier vers la droite, pour obtenir le rang de chaque année.

3. Construire le nuage de points représentant la série statistique double (x_i ; y_i) . Comparer cette représentation avec celle obtenue en **1.**.

4. a) On pose $z_i = y_i - 600$.

Indiquer la formule à saisir en **cellule B4**, et à recopier vers la droite, pour obtenir les valeurs z_i .

b) Construire le nuage de points représentant la série statistique (x_i ; z_i) .

5. Pour chaque commentaire suivant, indiquer le(s) nuage(s) de points qui le confirme(nt) :

a) la progression annuelle des nouveaux retraités est assez stable depuis 2004 ;

b) la progression annuelle des nouveaux retraités a fortement augmenté de 2005 à 2008.

6. a) En observant les trois graphiques, indiquer les nuages pour lesquels on peut envisager un ajustement affine.

b) Comparer les coefficients de corrélation linéaire des trois séries (a_i ; y_i) , (x_i ; y_i) et (x_i ; z_i) .

16 ★★ Le tableau ci-dessous compare les taux de chômage, exprimés en pourcentage, des jeunes de moins de 25 ans en Espagne et en Grèce, relevés au mois de janvier de chaque année :

	A	B	C	D
1	Année	Rang de l'année x_i	Taux en Espagne y_i	Taux en Grèce z_i
2	2008	1	20,7	22,7
3	2009	2	34,4	24,7
4	2010	3	30,8	30,2
5	2011	4	44,1	39,2
6	2012	5	50,7	52,1
7	2013	6	56,0	59,4
8	2014	7	54,1	57,1

1. À l'aide d'un tableur ou de la calculatrice, sur un même graphique, représenter les nuages de points des deux séries statistiques S1 (x_i ; y_i) et S2 (x_i ; z_i) .

2. a) Pour chaque nuage, leur forme permet-elle d'envisager un ajustement affine ?

b) Calculer le coefficient de corrélation linéaire de chacune des séries statistiques doubles S1 et S2.

Ces résultats permettent-ils de confirmer l'observation précédente ?

3. a) Pour chaque série S1 et S2, ajouter la droite de régression sur la représentation. Afficher leur équation.

b) À l'aide de ces ajustements, rechercher à partir de quelle année le taux de chômage des jeunes Grecs de moins de 25 ans deviendra supérieur à celui des jeunes Espagnols de moins de 25 ans.