

Ajustement affine

Point moyen									
Définitions et Notations	On étudie une population sur laquelle on observe deux caractères quantitatifs simultanément, notés x et y . On dispose donc, pour chaque individu, d'un couple de valeurs $(x_i; y_i)$								
	Exemple introductif. On mesure, pour 8 nageurs, leur taille (en cm) et leur envergure de bras (en cm) :								
	Nageur	1	2	3	4	5	6	7	8
	Taille x_i (cm)	165	170	172	178	180	183	188	190
	Envergure y_i (cm)	167	171	175	179	183	184	190	194
	nuage de points associé à cette série statistique double est l'ensemble des points $M_i(x_i; y_i)$ placés dans un repère.								
Exercice	Tracer dans un repère le nuage de points correspondant au tableau ci-dessus.								
Définition	Définition. On appelle point moyen du nuage le point $G(\bar{x}; \bar{y})$, où $\bar{x}$ et $\bar{y}$ sont les moyennes des séries (x_i) et (y_i) $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \ ; \ \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$								
Exemple	Reprenons les données précédentes : $\bar{x} = \frac{165 + 170 + 172 + 178 + 180 + 183 + 188 + 190}{8} = \frac{1426}{8} = 178,25$ $\bar{y} = \frac{167 + 171 + 175 + 179 + 183 + 184 + 190 + 194}{8} = \frac{1443}{8} = 180,375$ Donc : Donc $G(178,25 ; 180,375)$.								
Exercice	Placer dans le nuage de points précédents le point moyen								

Ajustement affine	
Définitions et Notations	Constat. Lorsque les points du nuage semblent globalement alignés (on parle d'une tendance linéaire), il est pertinent de chercher une droite qui « résume » cette tendance : c'est un ajustement affine du nuage de points. Méthode 1 — Ajustement visuel passant par G. Une droite d'ajustement raisonnable doit passer par le point moyen G et être tracée « au jugé » de façon à passer au plus près de l'ensemble des points, en laissant à peu près autant de points au-dessus qu'en dessous. Méthode 2 — Ajustement par la calculatrice ou le tableur (méthode des moindres carrés). En pratique, on utilise la fonction de régression linéaire de la calculatrice ou d'un tableur, qui calcule automatiquement l'équation de la droite $y = ax + b$ minimisant l'écart entre les points et la droite. Cette droite passe toujours par le point moyen G.

Exemple	Sur l'exemple, la calculatrice donne (arrondi) : $y \approx 1,09x - 13,9$
	On vérifie que G vérifie (approximativement) cette équation : $1,09 \times 178,25 - 13,9 \approx 180,4 \approx \bar{y} \checkmark$

Interpolation																						
Définitions et Notations	Une fois l'équation de la droite d'ajustement obtenue, on peut l'utiliser pour estimer une valeur de y à partir d'une valeur de x (ou inversement), sans avoir mesuré cette valeur. Définition — Interpolation. Estimer une valeur à l'intérieur de l'intervalle des données observées.																					
Exemple	Estimer l'envergure d'un nageur de 176 cm (valeur comprise entre 165 et 190, l'intervalle des tailles observées).* $y \approx 1,09 \times 176 - 13,9 \approx 178,0 \text{ cm}$																					
Définitions	Extrapolation. Estimer une valeur en dehors de l'intervalle des données observées.																					
Exemple	Estimer l'envergure d'un nageur de 200 cm (valeur en dehors de l'intervalle [165 ; 190]).* $y \approx 1,09 \times 200 - 13,9 \approx 204,1 \text{ cm}$																					
Remarque	Point de vigilance (essentiel). L'interpolation est en général fiable, car elle reste dans la zone où le modèle affine a été construit et validé par les données. L'extrapolation est plus risquée : rien ne garantit que la tendance linéaire observée se poursuive au-delà des valeurs mesurées. Plus on s'éloigne de l'intervalle des données, moins l'estimation est fiable.																					
Exercice	Une entreprise relève ses dépenses de publicité x (en milliers d'euros) et son chiffre d'affaires y (en milliers d'euros) sur 6 mois : <table><tr><th>Mois</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th></tr><tr><td>x_i</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>8</td></tr><tr><td>y_i</td><td>20</td><td>25</td><td>29</td><td>34</td><td>38</td><td>47</td></tr></table> - Tracer le nuage de points associé à cette série. - Calculer les coordonnées du point moyen G. - Déterminer, à la calculatrice, une équation de la droite d'ajustement affine. - En utilisant cet ajustement, estimer le chiffre d'affaires pour une dépense publicitaire de 5,5 milliers d'euros. S'agit-il d'une interpolation ou d'une extrapolation ? - Estimer le chiffre d'affaires pour une dépense de 15 milliers d'euros. S'agit-il d'une interpolation ou d'une extrapolation ? Ce résultat vous semble-t-il fiable ? Justifier.	Mois	1	2	3	4	5	6	x_i	2	3	4	5	6	8	y_i	20	25	29	34	38	47
Mois	1	2	3	4	5	6																
x_i	2	3	4	5	6	8																
y_i	20	25	29	34	38	47																