STATISTIQUES À DEUX VARIABLES

I. Série statistique à deux variables

1) Nuage de points

On considère deux variables statistiques et observées sur une même population de individus.
On note les valeurs relevées pour la variable et les valeurs relevées pour la variable .
Les couples forment une série statistique à deux variables.

Dans ce chapitre, on va s’intéresser au lien qui peut exister entre ces deux variables.

Définition : Dans un repère orthogonal, l’ensemble des points de coordonnées avec , est appelé le **nuage de points** associé à la série statistiques à deux variables.

 2) Point moyen

Définition : Le point G de coordonnées où et sont les moyennes respectives des et des , est appelé le **point moyen** du nuage de points associé à la série statistique à deux variables.

Méthode : Représenter un nuage de points

 **Vidéo** [**https://youtu.be/Nn6uckb3RvE**](https://youtu.be/Nn6uckb3RvE)

Le tableau suivant présente l’évolution du budget publicitaire et du chiffre d’affaire d’une société au cours des 6 dernières années :

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Budget publicitaire en milliers d’euros *xi* | 8 | 10 | 12 | 14 | 16 | 18 |
| Chiffre d’affaire en milliers d’euros *yi* | 40 | 55 | 55 | 70 | 75 | 95 |

1) Dans un repère, représenter le nuage de points (*xi*; *yi*).

2) Déterminer les coordonnées du point moyen G du nuage de points.

1)



2) = (8 + 10 + 12 + 14 + 16 + 18) : 6 = 13

 = (40 + 55 + 55 + 70 + 75 + 95) : 6 = 65.

Le point moyen G du nuage de points a pour coordonnées (13 ; 65). On peut placer ce point dans le repère.



II. Ajustement affine

 1) Interpolation, extrapolation

L’objectif est, à partir des valeurs d’une série statistique à deux variables, d’obtenir des approximations pour des valeurs inconnues de cette série.

Exemples :

- On donne une série exprimant la population d’une ville en fonction des années et on souhaite faire des prévisions pour les années à venir.
Les prévisions sortent du domaine d’étude de la série, on parle dans ce cas d’**extrapolation**.

- On donne une série exprimant la température extérieure et la consommation électrique correspondante. Les températures étudiées s’échelonnent entre -10°C et 10°C avec un pas de 4°C.

Sans faire de nouveaux relevés, on souhaite estimer la consommation électrique pour toutes les températures entières comprises entre -10°C et 10°C.

Les calculs sont dans le domaine d’étude de la série, on parle dans ce cas d’**interpolation**.

Définitions : L'**interpolation** et l’**extrapolation** sont des méthodes qui consistent à estimer une valeur inconnue dans une série statistique.

- Pour une interpolation, le calcul est réalisé dans le domaine d'étude fourni par les valeurs de la série.

- Pour une extrapolation, le calcul est réalisé en dehors du domaine d'étude.

La méthode d’extrapolation est parfois contestable car en dehors du domaine d’étude fourni par les valeurs de la série. Rien ne nous assure en effet que le modèle mathématique mis en œuvre soit encore valable.

 2) Droite d’ajustement

Pour obtenir de telles estimations, il faudra déterminer une droite passant « le plus près possible » des points du nuage.

L’interpolation ou l’extrapolation consistent à effectuer l’estimation par lecture graphique sur la droite ou par calcul à l’aide de l’équation de la droite.

Définition : Lorsque les points d’un nuage sont sensiblement alignés, on peut construire une droite, appelé **droite d’ajustement (ou droite de régression)**, passant « au plus près » de ces points.

Dans la suite, nous allons étudier différentes méthodes permettant d’obtenir une telle droite.

3) Méthode de Mayer

Cet ajustement consiste à déterminer la droite passant par deux points moyens du nuage de point.

Méthode : Déterminer la droite d’ajustement par la méthode des points moyens

 **Vidéo** [**https://youtu.be/ESHY4QPgriw**](https://youtu.be/ESHY4QPgriw)

On reprend les données de la méthode du paragraphe I.

1) Soit G1, le point moyen associé aux trois premiers points du nuage et G2 le point moyen associé aux trois derniers points du nuage.

 a) Calculer les coordonnées de G1 et G2.

 b) On prend (G1G2) comme droite d’ajustement. Tracer cette droite.

2) À l’aide du graphique :

 a) Estimer le chiffre d’affaire à prévoir pour un budget publicitaire de 22 000 €.

 b) Estimer le budget publicitaire qu’il faudrait prévoir pour obtenir un chiffre d’affaire de 100 000 €.

 c) La méthode utilisée dans les questions 2a et 2b consiste-t-elle en une interpolation ou une extrapolation ?

1) a) = (8 + 10 + 12) : 3 = 10

 = (40 + 55 + 55) : 3 = 50.

Le point moyen G1 a pour coordonnées (10 ; 50).

 = (14 + 16 + 18) : 3 = 16

 = (70 + 75 + 95) : 3 = 80.

Le point moyen G2 a pour coordonnées (16 ; 80).

 b)



2) On lit graphiquement :

 a) Le chiffre d’affaire à prévoir pour un budget publicitaire de 22 000 € est de

110 000 €.

 b) Le budget publicitaire qu’il faudrait prévoir pour obtenir un chiffre d’affaire de

100 000 € est de 20 000€.



c) Les lectures graphiques sont réalisées ici en dehors du domaine d’étude, on parle donc d’extrapolation.

4) Méthode des moindres carrés

Cette méthode porte le nom de « moindre carrés » car elle consiste à rechercher la position de la droite d’ajustement tel que la somme des carrés des longueurs donnant les distances respectives (en vert) entre la droite et les points soit minimale.

Le principe consiste donc à déterminer les coefficients et d’une droite d’équation de sorte qu’elle passe le « plus près possible » des points du nuage.

Pour chaque abscisse , on calcule la distance entre
le point du nuage et le point de la droite, soit :

Il s’agit dans ce cas, de la droite d’ajustement de en .

A noter : Il existe également une droite d’ajustement de en en calculant les distances obtenues par projection horizontale.



Dans la méthode des moindres carrés, on recherche et pour lesquels la somme des carrés des distances est minimale, soit :

 est minimale.

Pour cela, on peut appliquer la propriété suivante :

Propriété : La droite d’ajustement de en a pour équation , avec :

où est la covariance de

et est la variance de .

*- Admis -*

Méthode : Déterminer la droite d’ajustement par la méthode des moindres carrés

 **Vidéo** [**https://youtu.be/vdEL0MOKAIg**](https://youtu.be/vdEL0MOKAIg)

On considère la série statistique à deux variables données dans le tableau suivant :

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *xi* | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 |
| *yi* | 13 | 23 | 34 | 44 | 50 | 65 | 75 | 90 |

1) Dans un repère, représenter le nuage de points (*xi*; *yi*).

2) a) Déterminer une équation de la droite d’ajustement par la méthode des moindres carrés.

 b) Vérifier à l’aide de la calculatrice.

 c) Représenter la droite d’ajustement de *y* en *x*.

3) Estimer graphiquement la valeur de *x* pour *y* = 70. Retrouver ce résultat par calcul.

S’agit-il d’une interpolation ou d’une extrapolation ?

1)



2) a) On commence par calculer, les moyennes et

Par la méthode des moindres carrés, la droite d’ajustement de en a pour équation avec :

Et

Une équation de la droite d’ajustement est :

Pour le tracé, on considère l’équation :

b) Avec TI :

- Appuyer sur « **STAT**» puis « **Edite**» et saisir les valeurs de *xi* dans L1 et les valeurs de *yi* dans L2.

- Appuyer à nouveau sur « **STAT**» puis « **CALC**» et « **RegLin(ax+b)**».

- Saisir **L1,L2**

Avec CASIO :

- Aller dans le menu « **STAT**».

- Saisir les valeurs de *xi* dans *List1* et les valeurs de *yi* dans *List2*.

- Sélectionner « **CALC**» puis « **SET**».

- Choisir *List1* pour *2Var XList* et *List2* pour *2Var YList* puis « **EXE**».

- Sélectionner « **REG** » puis « **X**» et « **aX+b**».

La calculatrice nous renvoie : et

Une équation de la droite d’ajustement est :

Pour tracer la droite, il suffit de calculer les coordonnées de deux points de la droite d’ajustement :

- Si alors donc le point de coordonnées appartient à la droite d’ajustement.

- Si alors donc le point de coordonnées appartient à la droite d’ajustement.



c)

3) - Pour , on lit graphiquement.

 - Par calcul, si , alors

Soit

 - Les calculs sont réalisés dans domaine d’étude, on parle donc d’interpolation.

5) Coefficient de corrélation

Définition : Le **coefficient de corrélation** de et est donné par :

Interprétation :

Le coefficient de corrélation est un nombre compris entre -1 et 1 qui mesure la relation entre les deux variables et . Plus le coefficient est proche des valeurs extrêmes -1 et 1, plus la corrélation linéaire entre les variables est forte.

- Si , les valeurs prises par ont tendance à croître quand les valeurs de augmentent.

- Si , les valeurs prises par ont tendance à décroître quand les valeurs de augmentent.

- Si , les variations des variables et sont indépendantes.

Exemples de coefficients de corrélation :



Méthode : Calculer un coefficient de corrélation

 **Vidéo** [**https://youtu.be/FxREenh3fgE**](https://youtu.be/FxREenh3fgE)

En reprenant les données de la méthode précédente, calculer le coefficient de corrélation et interpréter le résultat.

Soit :

Le coefficient de corrélation est proche de 1 donc la corrélation entre les deux variables est fortes. Les points du nuage sont proches de la droite d’ajustement.

III. Ajustement par changement de variable

Lorsque le nuage de points n'est à priori pas modélisable par une droite, on peut réaliser un ajustement linéaire en effectuant un changement de variable.

Méthode : Effectuer un ajustement se ramenant par changement de variable à un ajustement affine

 **Vidéo** [**https://youtu.be/nVDL0razClY**](https://youtu.be/nVDL0razClY)

On a relevé la population d’une grande métropole sur 50 ans tous les 5 ans. Les résultats sont présentés dans le tableau suivant :

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Année  | 0 | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 |
| Populationen milliers  | 19,4 | 19,4 | 27,6 | 40,3 | 50 | 59 | 69 | 87 | 132 | 166 | 216 |

1) Représenter le nuage de points dans un repère.

2) a) On effectue le changement de variable . Réaliser un nouveau tableau présentant les valeurs prises par les variables et .

 b) Représenter un nouveau nuage de points à partir des données des variables et .

 c) A l’aide la calculatrice, déterminer une équation de la droite d’ajustement de *z* en *x* par la méthode des moindres carrés. Représenter la droite d’ajustement.

3) a) En déduire la relation qui lie et puis tracer la courbe représentative de la fonction définie par dans le repère contenant le premier nuage de points.

 b) En admettant que le modèle mathématique reste valable en dehors du domaine d’étude, extrapoler le nombre d’habitant 5 ans après l’étude.

1)

2) a)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 |
|  | 3 | 3 | 3,3 | 3,7 | 3,9 | 4,1 | 4,2 | 4,5 | 4,9 | 5,1 | 5,4 |

b)



c) Une équation de la droite d’ajustement est :

On trace la droite : voir ci-dessus.



3) On a :

 et : , soit :

Donc :

 est l’expression de la fonction permettant d’ajuster le nuage de points .

b)

On peut supposer que 5 années après la fin de l’étude, la population de la ville sera proche de 276 000 habitants.

Hors du cadre de la classe, aucune reproduction, même partielle, autres que celles prévues à l'article L 122-5 du code de la propriété intellectuelle, ne peut être faite de ce site sans l'autorisation expresse de l'auteur.

[*www.maths-et-tiques.fr/index.php/mentions-legales*](http://www.maths-et-tiques.fr/index.php/mentions-legales)