Statistiques à deux variables

Première STMG et STI2D · commun · Chapitre 05

Représenter deux variables par un nuage

Définition · Comprendre la notion

Une série à deux variables quantitatives contient des couples $(x_i ; y_i)$ observés sur les mêmes individus. Un nuage de points place chaque couple dans un repère. Annonce les variables, leurs unités et l'échelle des axes. Intervertir les colonnes modifie la question étudiée : prédire y à partir de x n'est pas la même opération que prédire x à partir de y. Le nuage permet d'observer une tendance, une courbure, des groupes ou des valeurs atypiques. Une tendance croissante ne signifie pas que chaque point est au-dessus du précédent ni qu'une relation de cause à effet est démontrée.

Le point moyen du nuage

Définition · Comprendre la notion

Pour n couples, le point moyen est $G(\bar{x} ; \bar{y})$, avec $\bar{x} = (x_1 + \cdots + x_n)/n$ et $\bar{y} = (y_1 + \cdots + y_n)/n$. Calcule séparément les moyennes de chaque coordonnée, en gardant leurs unités. Le point moyen résume le centre du nuage, mais ne suffit pas à décrire sa forme. La droite de régression affine des moindres carrés avec constante passe par G. Une droite choisie au jugé ne possède pas automatiquement cette propriété. Si une droite de pente a doit passer par G, son ordonnée à l'origine est $b = \bar{y} - a\bar{x}$.

Construire un ajustement affine

Définition · Comprendre la notion

Un ajustement remplace un nuage imparfaitement aligné par une relation approchée $y = ax + b$. Une droite au jugé, une droite de Mayer ou une régression peuvent donner des ajustements différents. La méthode de Mayer partage les points ordonnés en deux groupes et relie leurs points moyens. La méthode des moindres carrés choisit a, b pour minimiser la somme des carrés des écarts verticaux $\sum (y_i - (ax_i + b))^2$; on utilise un outil numérique pour la déterminer. L'écart $y_i - \hat{y}_i$ s'appelle un résidu. Un nuage courbé ou des résidus organisés signalent qu'une droite peut être inadéquate. Arrondis les coefficients seulement à la précision utile.

Interpolation et extrapolation

Définition · Comprendre la notion

Une interpolation estime une valeur pour une abscisse à l'intérieur de l'intervalle observé. Une extrapolation utilise le modèle en dehors de cet intervalle ; elle demande davantage de prudence. Pour prévoir y, remplace x dans le modèle. Pour chercher la valeur de x correspondant à un objectif, résous l'équation du modèle. Si x est une date codée, reviens ensuite à la date réelle. Même à l'intérieur de l'intervalle, une prévision n'est pas une certitude individuelle. Vérifie son ordre de grandeur, son unité et la pertinence du modèle ; évite les extrapolations incompatibles avec une limite physique ou un marché saturé.

Association, causalité et qualité d'ajustement

> **Définition · Comprendre la notion**
>
> Deux variables peuvent évoluer ensemble sans que l'une cause directement l'autre. Une troisième variable, un effet de sélection ou une coïncidence peuvent expliquer l'association. Pour une régression affine avec constante, le coefficient de détermination R^2 mesure la part de variation des ordonnées expliquée par le modèle sur les données. Une valeur proche de 1 traduit un bon ajustement au sens de ce critère, mais ne prouve ni causalité ni validité future. Regarde toujours le nuage, le nombre d'observations, les résidus et les conditions de recueil. Un point extrême peut influencer fortement une régression. Des populations différentes peuvent masquer des relations différentes.

Les repères essentiels

Représenter deux variables par un nuage.

1. Identifie les deux variables et garde chaque paire de mesures associée.
2. Place x horizontalement et y verticalement, avec des axes nommés et gradués.
3. Décris la tendance, la courbure, les groupes et les points atypiques.
4. Compare des taux de variation si tu recherches un alignement exact.
5. Vérifie les données atypiques sans les supprimer uniquement pour améliorer l'apparence.

Le point moyen du nuage.

1. Compte le nombre de couples et calcule la moyenne des abscisses.
2. Calcule séparément la moyenne des ordonnées avec le même nombre d'observations.
3. Écris G avec les coordonnées dans le bon ordre et leurs unités.
4. Pour une droite passant par G, remplace x et y par ces deux moyennes.
5. Contrôle le passage par G et ne déduis pas la forme du nuage de son seul centre.

Construire un ajustement affine.

1. Observe le nuage et précise la méthode d'ajustement utilisée.
2. Calcule ou relève les coefficients a et b avec une précision suffisante.
3. Calcule les prévisions $\hat{y}_i$, puis les résidus $y_i - \hat{y}_i$.
4. Compare les carrés des résidus et leur organisation, pas seulement leur somme signée.
5. Présente les valeurs obtenues comme des prévisions dépendant du modèle.

Interpolation et extrapolation.

1. Relève l'intervalle observé, le codage de x et les unités.
2. Convertis la demande en abscisse et distingue interpolation et extrapolation.
3. Calcule une image ou résous l'équation correspondant à l'objectif.
4. Reconvertis le résultat en date ou en quantité et arrondis avec une précision justifiée.
5. Contrôle les limites physiques et indique la dépendance au modèle.

Association, causalité et qualité d'ajustement.

1. Précise si la question concerne l'ajustement, la prévision ou une explication causale.
2. Compare les indicateurs uniquement sur des données et une variable comparables.
3. Examine le nuage, les résidus, les valeurs atypiques et la taille de la série.
4. Cherche une cause commune, une sélection ou une causalité inverse avant de conclure.

5. Limite la conclusion à ce que le protocole et les données permettent réellement d'affirmer.

Un exemple pour se repérer

Exemple · Exemple commenté

Pour les points $(0;1)$, $(1;3)$, $(2;4)$, une régression affine donne $\hat{y} = 1{,}5x + 7/6$. Les prévisions sont $7/6$, $8/3$ et $25/6$. Les résidus observation moins prévision sont donc $-1/6$, $1/3$ et $-1/6$.

Leur somme est nulle, mais les erreurs ne sont pas toutes nulles. La somme des carrés vaut $1/36 + 1/9 + 1/36 = 1/6$. Ce calcul rend visible une dispersion que la somme signée annulerait.

Le point moyen est $G(1;8/3)$. La droite passe bien par lui puisque $1{,}5 + 7/6 = 8/3$. Pour un affichage, on peut arrondir la constante à 1,1667 ; pour contrôler les résidus, garder $7/6$ évite les petites erreurs d'arrondi.

Les erreurs à éviter

Attention · Comprendre une erreur fréquente

Une observation atypique peut être vraie. Avant d'utiliser le mot « erreur », distingue une erreur de saisie vérifiée d'une mesure inhabituelle mais réelle.