Statistiques à deux variables BTS

ZEG

# STATISTIQUE À DEUX VARIABLES - COURS

## 1. Présentation de la situation

Une étude statistique est réalisée sur le montant des ventes d’un chalutier au fil des semaines. Deux caractères sont relevés simultanément : - Le rang de la semaine - Le montant de la vente de la semaine (en milliers d’euros)

**Tableau des données :**

| Semaines | S1 | S2 | S3 | S4 | S5 | S6 |
| --- | --- | --- | --- | --- | --- | --- |
| Rang | 1 | 2 | 3 | 4 | 5 | 6 |
| Montant | 23 | 19 | 33 | 27 | 38 | 40 |

**Problématique :** - Existe-t-il un lien entre ces deux caractères ? - Peut-on donner une estimation du montant des ventes de la 14ème semaine ?

## 2. Série statistique à 2 variables

Définition 1 Une série statistique à deux variables est une série dans laquelle on relève pour chaque individu deux caractères mesurables.

**Remarque 1** L’intérêt d’une telle étude réside dans la recherche d’un lien éventuel entre les deux caractères étudiés.

## 3. Représentation graphique

**Définition 2** Dans un repère orthogonal, l’ensemble des points de coordonnées constitue un **nuage de points**.

**Exemple 1** - Nuage de points représentant la situation étudiée :

![Nuage de points](data:binary/octet-stream;base64,)

Nuage de points

**Remarque 2** Ce type de représentation donne un aperçu de la dispersion des données.

## 4. Point moyen

**Définition 3** Le point moyen d’un nuage de points est le point de coordonnées où : - est la moyenne des abscisses - est la moyenne des ordonnées

**Exemple 2** - Dans la situation étudiée :

![Point moyen G](data:binary/octet-stream;base64,)

Point moyen

## 5. Ajustement affine

Réaliser un ajustement affine consiste à déterminer une droite qui passe le plus près possible de tous les points du nuage.

### 5.1 Ajustement affine à la règle

Méthode utilisée pour des lectures graphiques (interpolations ou extrapolations).

* **Interpolation** : Évaluation d’une variable dans les limites de l’échantillon
* **Extrapolation** : Évaluation d’une variable hors des limites de l’échantillon

![Interpolations ou extrapolations](data:binary/octet-stream;base64,)

Interpolations ou extrapolations

### 5.2 Ajustement affine par la droite de Mayer

Cette méthode consiste à diviser le nuage de points en deux sous-nuages de même importance.

Propriété 1 La droite de Mayer passe par , le point moyen du nuage.

![Droite de Mayer](data:binary/octet-stream;base64,)

Droite de Mayer

### 5.3 Ajustement affine par la méthode des moindres carrés

Soit un nuage de points . On considère une fonction affine du type . Pour chaque point , on calcule :

![Méthode des moindres carrés](data:binary/octet-stream;base64,)

Méthode des moindres carrés

**Propriété 2** Il existe une unique droite telle que la somme des carrés des écarts soit minimale. Cette droite, appelée **droite de régression de en** , passe par .

## 6. Coefficient de corrélation linéaire

Définition 4 La qualité d’un ajustement affine s’évalue à l’aide d’un nombre , appelé le coefficient de corrélation linéaire.

Propriété 3 - - Les points du nuage sont alignés si et seulement si ou - a le même signe que le coefficient directeur de la droite de régression - Plus est proche de ou , meilleur est l’ajustement - Si , il n’est pas conseillé de réaliser un ajustement affine

**Exemple 6** - Application à notre situation :

donc un ajustement affine est envisageable. La droite de régression de en a pour équation réduite :

**Réponse à la problématique :** Estimation pour la 14ème semaine : milliers d’euros

## 7. Commandes Excel utiles

* Pour : =INDEX(DROITEREG(Y;X);1)
* Pour : =INDEX(DROITEREG(Y;X);2)
* Pour : =COEFFICIENT.DETERMINATION(Y;X)

## 8. Relation de cause à effet

**Attention :** Un coefficient de corrélation élevé n’induît pas forcément une relation de causalité entre deux variables.

**Exemple 7** Le nombre de coups de soleil observés sur une plage peut être fortement corrélé au nombre de lunettes de soleil portées, mais aucun de ces deux phénomènes n’est la cause de l’autre !