



Modèle Linéaire

Aspects
opérationnels

Contexte
théorique

Echantillonnage

Planification

Situations

Sélection

Démarche

Deux phases

Modèle Linéaire - Régression

Aspects opérationnels

Thierry Dhorne

Institut Universitaire de Technologie de Vannes
Université de Bretagne Sud

Année Universitaire 2014-2015



Contexte théorique de la régression

Modèle Linéaire

Aspects
opérationnels

Contexte
théorique

Echantillonnage

Planification

Situations

Sélection

Démarche

Deux phases

- il existe 2 visions théoriques de la régression linéaire
- ➔ qui recouvrent 2 contextes expérimentaux différents
 - ▶ l'échantillonnage,
 - ▶ la planification.



Échantillonnage

Modèle Linéaire

Aspects
opérationnels

Contexte
théorique

Échantillonnage

Planification

Situations

Sélection

Démarche

Deux phases

- on tire un couple de variables aléatoires (Y, X) puis on mesure des valeurs de X et de Y
- c'est l'exemple des moules vu précédemment
- le modèle de régression peut alors être considéré (d'un point de vue théorique) comme la spécification de Y conditionnellement à $X = x$
- pas de différence avec le modèle de référence dès lors que l'on raisonne à $X = x$ connu (variable aléatoire $Y/X = x$)
- on décompose la loi du couple (X, Y) en loi de X et loi de $Y/X = x$



Planification

Modèle Linéaire

Aspects
opérationnels

Contexte
théorique

Echantillonnage

Planification

Situations

Sélection

Démarche

Deux phases

- on choisit les valeurs de la variable explicative pour lesquelles on veut mesurer la variable à expliquer
- ➔ exemple de l'estimation locale de la constante de gravitation g
- ➔ la variable explicative est considérée comme fixée (non aléatoire)
- seule la variable expliquée est aléatoire



Estimation de l'accélération de la pesanteur

Modèle Linéaire

Aspects
opérationnels

Contexte
théorique

Echantillonnage

Planification

Situations

Sélection

Démarche

Deux phases

$$k(l - l_0) = F.$$

où k est la constante de raideur du ressort

$$l = l_0 + \frac{g}{k}m,$$

où m est la masse suspendue au ressort

- on réalise des mesures de la longueur pour $n(\geq 2)$ masses différentes
- ➔ les masses sont choisies (non aléatoires)
- ➔ il y a une incertitude sur la longueur (\Rightarrow plusieurs mesures de longueur)
- d'où le modèle

$$L_r = \lambda_0 + \kappa m_r + E_r$$



Situations réelles

Modèle Linéaire

Aspects
opérationnels

Contexte
théorique

Echantillonnage

Planification

Situations

Sélection

Démarche

Deux phases

- l'échantillonnage (parfois exhaustif) correspond à la situation du Data Mining

→ je dispose d'un fichiers de clients dans lequel je cherche des variables qui expliquent une variable donnée

- la planification correspond à la situation de la statistique expérimentaliste (par exemple biostatistique)

→ je choisis les doses de médicaments que j'administre à différents malades pour connaître l'influence du médicament sur une variable physiologique



Sélection

Modèle Linéaire

Aspects
opérationnels

Contexte
théorique

Echantillonnage

Planification

Situations

Sélection

Démarche

Deux phases

- il existe une situation intermédiaire entre les deux cas précédents
- ➔ lorsque l'on échantillonne le couple (Y, X) (cas précédent), on peut décider de ne mesurer Y que pour certaines valeurs de $X = x$
- ➔ pour optimiser le rapport précision / coût expérimental (coût d'obtention de la variable d'intérêt important)
- par exemple, dans le fichier de clients, je ne choisis que les très jeunes et les très âgés



La démarche globale

Modèle Linéaire

Aspects
opérationnels

Contexte
théorique

Echantillonnage

Planification

Situations

Sélection

Démarche

Deux phases

1. identifier un problème de prédiction,
2. identifier la variable à prédire,
3. identifier la variable prédictrice,
4. vérifier que les deux variables sont quantitatives,
5. vérifier que le lien est linéaire,
6. vérifier que la variance est constante,
7. vérifier que les «observations» sont indépendantes,



Les deux phases

Modèle Linéaire

Aspects
opérationnels

Contexte
théorique

Echantillonnage

Planification

Situations

Sélection

Démarche

Deux phases

- explication
- étude du lien entre les deux variables et modélisation
- mise au point du modèle
- prédiction
- utilisation du modèle pour prédire la valeur de Y inconnue pour une nouvelle valeur de x connue