

# Modélisation des temps de vente de véhicules d'occasion.

Solohaja-Faniaha Dimby

Université de Paris I  
90 rue de Tolbiac, Paris, France,  
Solohaja-Faniaha.Dimby@malix.univ-paris1.fr

Modèles et Apprentissage en Sciences Humaines et Sociales  
Paris, 04 - 05 Juin 2012

# Outline

Modélisation des  
temps de vente de  
véhicules  
d'occasion.

Solohaja-Faniaha  
Dimby

Motivation

Présentation

Les données

Méthodologie

Conclusions et  
perspectives

- 1 Motivation
- 2 Présentation
- 3 Les données
- 4 Méthodologie
- 5 Conclusions et perspectives

## Motivation

Présentation

Les données

Méthodologie

Conclusions et  
perspectives

- L'étude fait suite à un travail sur la modélisation et prédiction des prix de VO (véhicules d'occasion).
- Effectuée à la demande des experts du marché automobile.
- Politique des constructeurs : accélération de la rotation du parc automobile (acheter des véhicules plus souvent).
- Le marché de l'occasion représente un enjeux considérable.
- Marché du neuf est lié fortement à celui de l'occasion.

- Il s'agit de modéliser le temps de vente de VO.
- Deux approches possibles et complémentaires :
  - Approche qualitative :description et test d'association.
  - Approche quantitative : modèle de prédiction.

# Description de la base de données

Modélisation des  
temps de vente de  
véhicules  
d'occasion.

Solohaja-Faniaha  
Dimby

Motivation

Présentation

**Les données**

Méthodologie

Conclusions et  
perspectives

- VO mis en vente sur internet sur la période de décembre 2010 à février 2012.  
Après nettoyage, 2 388 600 individus.
- Données sur les caractéristiques de chaque véhicule : marque, modèle, version, carrosserie, carburant, prix (proposé à la vente), âge, km, âge (au moment de la mise en vente).
- Données sur le temps de vente : date de mise en ligne de l'annonce (première parution sur le web - méthodes de détection des doublons), date à laquelle l'annonce a été observée pour la dernière fois, date à laquelle le prix a été modifié pour la dernière fois.

Hypothèse : un véhicule est vendu lorsque l'annonce a été retirée.

# Les variables d'intérêts

Modélisation des  
temps de vente de  
véhicules  
d'occasion.

Solohaja-Faniaha  
Dimby

Motivation

Présentation

Les données

Méthodologie

Conclusions et  
perspectives

- Deux variables d'intérêts :  
 $tv$  = date de retrait - date d'apparition.  
 $tvAjust$  = date de retrait - date de changement de prix.
- Autre variables utilisées :  
 $km$ ,  
 $agefin = age + tv$ ,  
 $N$  = Nb de VO vendu par jour,  
 $t0 = tvAjust - tv$  (peut être important pour la suite),  $modif=1$  si le prix a été modifié et 0 sinon  
 $r = (prix - prix\ predit)/prix$ , prediction faite par modèle validé antérieurement.

## Distribution de la variable tvAjust : HISTORAMME ET STATISTIQUE DESCRIPTIVE

- Discrétisation de la variable tsAjust en 5 classes : très rapide, rapide, moyen, lent, très lent. Exemple :
- Discrétisation de la variable agefin :
  - classAge=1 si  $\text{agefin} < 12$  ;
  - classAge=2 si  $(12 \leq \text{agefin} \leq 36)$  ;
  - classAge=3 si  $(36 < \text{agefin} \leq 60)$  ;
  - classAge=4 si  $\text{agefin} > 60$  ;

Choix justifié par une étude sur les outliers, évolution de l'utilisation du véhicules. (voir graphique).

- Niveau d'agrégation des données : sur conseil des experts, vérifiés par l'analyse de la variance : modèle-carburant-carrosserie.

- Il s'agit de faire un typologie sur la rotation de vente et des test d'association.
  - regression polytomique, effectué par Proc logistic de SAS.
- (résultats sur fichier séparé).



Construction de modèle qui permet de prédire dans un intervalle, le temps de vente d'un VO proposé.

- Modèle linéaire : (transformation en log pour  $ts_{Ajust}$ , pour s'assurer de la positivité du temps prédit.)  
 $lts_{Ajust} = f(r, \text{modif}, \text{km}, \text{prix})$ .  
manque de précision, résultat dans un intervalle.

Exemple de résultats sur fichier séparé.

- La variable  $r$  est pertinente dans la détermination du temps de vente.
- Elle contient une information sur le prix.
- Le temps de vente est variable selon le modèle, carburant, carrosserie, tranche d'âge et le kilométrage.
- résultats concordants avec la réalité. (avis d'experts).

- En cours : modélisation par une loi Gamma , paramètre fonction de l'age.
- Voir le problème comme un processus de Poisson, en considérant le nombre de vente dans une période définie.