

Exemple de modélisation d'un prix de vente dans les services

Contexte: Agence immobilière qui possède un fichier de maisons vendues avec différents critères (type d'agence, superficie, nbr de pièces, emplacement etc...) sur l'année 2009

- Questions:
- Comment déterminer au mieux le prix de vente des biens pour les vendre plus vite et au bon prix?
 - Certains biens ne se vendent pas, est ce une sur-estimation du prix?
 - Suis-je, en terme de prix, bien placé par rapport à la concurrence?



Outils et méthodes utilisés:

- Régression linéaire multiple mixte (sur données continues et catégoriques)
- Analyse de résidus
- Test de Student

Analyse du fichier client

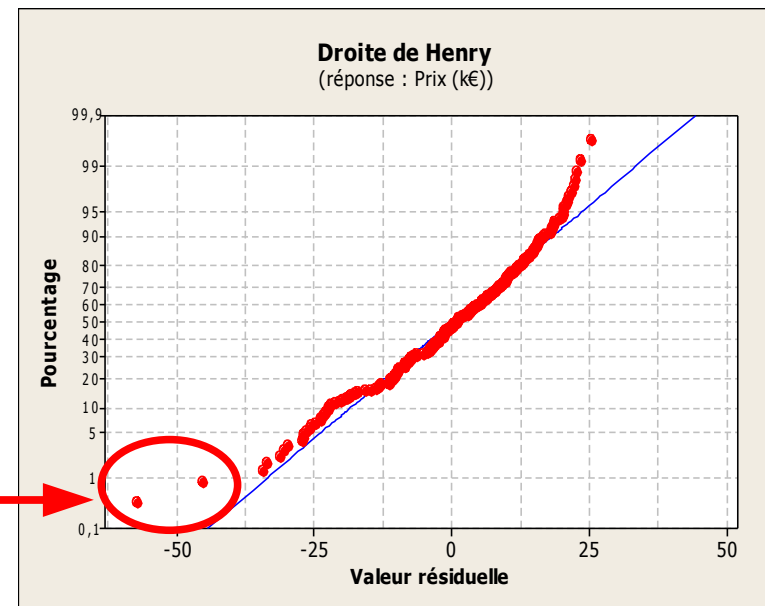


- Recherche de valeurs aberrantes

Observation	nbr chambre	Prix (k€)	Valeur ajustée	Ajust ErT	Valeur résiduelle	Valeur résiduelle normalisée
30	4,00	92,000	123,131	2,018	-31,131	-2,17R
53	4,00	70,000	115,119	3,319	-45,119	-3,19R
62	3,00	73,000	107,133	2,446	-34,133	-2,39R
84	4,00	100,000	156,918	3,072	-56,918	-4,01R
163	3,00	86,000	116,396	1,356	-30,396	-2,10R
173	3,00	62,000	95,469	2,977	-33,469	-2,36R
199	3,00	90,000	119,671	1,400	-29,671	-2,05R

Résultats:

- Différence d'estimation du bien de l'ordre de **50k€** (différence anormalement élevée et complètement en dehors de l'erreur d'estimation standard)
- Ces deux maisons vendues ont été sous-évaluées (bon plan pour l'acquéreur, nettement moins bien pour le vendeur)
- 1 des 2 sous-évaluations a été faite par notre agence



Droite qui représente pour chaque maison la différence entre l'estimation réelle et l'estimation du modèle

- Modélisation du prix de vente

L'équation de régression est

$$\text{Prix (k€)} = 12,0 + 0,774 \text{ surface(m}^2\text{)} + 23,1 \text{ zone géo_centre ville} - 6,68 \text{ zone géo_centre ville+15kms}$$

S = 7,59140 R carré = 87,9 % R carré (ajust) = 87,7 %

Analyse de variance

Source	DL	Somme des carrés	CM	F	P
Régression	3	87134	29045	503,99	0,000
Erreur résiduelle	209	12045	58		
Total	212	99178			

Résultats:

- Seules la superficie et la zone géographique rentrent dans l'estimation du prix des maisons. Un facteur comme l'ancienneté n'a pas d'influence.
- Notre agence n'évalue pas différemment les biens que les autres agences. Le prix de vente n'est donc pas un facteur de différenciation sur notre marché.
- Le modèle qui relie le prix de vente à la superficie et à la zone géographique est pertinent et réel => le modèle est bon





- Pourquoi certaines maisons ne se vendent pas?

Une maison ayant les caractéristiques suivantes ne se vend pas , est ce que le prix est trop élevé?

nbr chambre	ancienneté	zone géo	Agence	surface(m ²)	Prix (k€)
6	2	Centre ville+15kms	Nous	149	145

Valeurs prévues pour cette maison

Nouvelle observation	Valeur ajustée	Ajust ErT	IP à 95 %
1	120,613	1,268	(105,440; 135,786)

Résultat:

- Le modèle prévoit une valeur de l'ordre de 121k€ avec une marge d'erreur comprise entre 105k€ en estimation basse et 135k€ en estimation haute.
- Estimée à 145k€, ce bien est clairement sur estimé (10k€ au dessus de l'estimation haute du modèle).
- La raison de la difficulté à vendre ce bien est sans doute la mauvaise estimation du prix de vente.

Résultats de l'étude:

- Possibilité de déterminer un prix de vente à partir de seulement quelques critères clés (modélisation du prix de vente)et ceci de manière très fiable. Détermination du juste prix immédiatement (gain de temps sur la vente, pas de biens en souffrance,...).
- Certains facteurs n'influencent aucunement le prix de vente.
- La différenciation entre les agences ne se fait pas sur le prix de vente.
- Possibilité de savoir si certains biens sont sur ou sous-estimés.
- Possibilité, en fonction du marché de l'immobilier, d'adapter en permanence le modèle de prix de vente (voire de le complexifier avec les temps de vente , la marge , période de l'année etc...).
- Possibilité d'identifier des marchés de niches où les critères de prix sont différents.
- ...

