



ETUDE SUR LE PRIX DES VINS DE BOURGOGNE

Jean-Sébastien Wenger

MODELAB Laboratoire de modèles mathématiques

196 rue de Bernex

1233 Bernex

Genève

Suisse

Tél. et fax: Suisse: 022 777 20 23, autres pays: 4122 777 20 23

E-mail: taxigeneve@bluewin.ch

Résumé

Cette étude a pour but d'identifier quelles sont les variables les plus significatives qui influencent la formation du prix des vins de Bourgogne. Elle est basée sur les éléments fournis par le guide GaultMillau.

Le résultat démontre que le prix répond à des critères qualitatifs et de coûts.

Introduction

Le guide des vins GaultMillau, édition 2004, fournit une série de données sur les domaines visités. Pour les producteurs qui l'ont donné, il indique aussi le prix du vin. Nous avons tenté de voir si l'on peut expliquer le prix à l'aide de cette série de données.

Les variables

N'ont été introduits dans la base de données que les vins qui ont le prix affiché et les domaines qui présentent toutes les variables explicatives présent en compte. Ces variables sont au nombre de quinze. Elles sont:

1) la sous région

Classées du Nord au Sud, les sous régions sont numérotées:

41 pour le Chablis,

42 pour la Côte de Nuits,

43 pour la Côte de Beaune,

44 pour la Côte Chalonnaise et

45 pour le Mâconnais.

(Le 4 provient du numéro donné à la Bourgogne, selon le même critère.)

(Le Beaujolais ne fait pas partie de cette statistique.)

2) vigneron indépendant

Cette variable est notée 1 s'il s'agit d'un vigneron indépendant et 0 si ce n'est pas le cas.

3) cotation du domaine

La variable correspond au nombre d'étoiles, avec 0,5 pour une étoile entre parenthèses.



4) couleur du vin en question

Cette variable prend la valeur
1 pour un vin blanc,
2 pour un vin rosé,
3 pour un vin rouge.

5) surface du vignoble

Le nombre d'hectares indiqués dans le guide.

6) surface du vignoble de la couleur du vin en question

Pour le rosé, il s'agit de la surface totale du vignoble, soit la même que la variable précédente.

7) âge moyen des vignes

Quand plusieurs âges sont mentionnés, il a été calculé la moyenne, arrondie à l'unité.

8) vendange manuelle

Cette variable est notée 100 pour une vendange totalement manuelle et 0 pour une vendange totalement mécanique, sinon proportionnellement.

9) production moyenne

La production est notée en kilo bouteille, c'est-à-dire en mille bouteilles.

10) coup de coeur GaultMillau

Cette variable prend la valeur 1 si le vin correspond à un coup de coeur GaultMillau et 0 si ce n'est pas le cas.

11) millésime

L'année de production du vin.

12) cotation du millésime

Les valeurs utilisées sont les cotations données par le guide à la page 28, la mention « excellent » étant remplacée par le chiffre 5.

13) cotation GaultMillau

La valeur donnée dans le guide.

14) cru

Cette variable prend la valeur
2 pour un grand cru,
1 pour un 1^{er} cru,
0 pour les autres vins.

15) garde

Cette variable prend la valeur



- 1 pour un vin à boire tout de suite,
- 2 pour un vin à boire tout de suite ou à garder,
- 3 pour un vin à garder.

La variable dépendante est donc le prix du vin indiqué en euros.

La base de données

La base de données ainsi constituée représente 631 vins, dont 247 vins blancs, 1 vin rosé et 383 vins rouges.

Il y a 297 vins sans mentions de premier cru ou grand cru, 260 mentionnés "premier cru" et 74 vins mentionnés "grand cru".

Rappelons ici que le cru est un critère géographique de qualité, c'est-à-dire qu'il s'agit de terroirs dont les vins ont été reconnus de grande qualité.

Le modèle

Le meilleur modèle auquel nous soyons parvenus est le suivant, qui explique le prix à l'aide de 9 variables:

$$\hat{p} = -287.08 + 2.9065 \cdot \text{vigind} + 1.3182 \cdot \text{coul} + 0.13946 \cdot \text{ha} - 0.14602 \cdot \text{hacoul} + 0.14483 \cdot \text{age} + 0.085635 \cdot \text{vendmanu} + 0.013967 \cdot \text{prod} + 3.2185 \cdot \text{cot gm} + 11.098 \cdot \text{cru}$$

où:

vigind = vigneron indépendant; $t_{stat} = 2.324$

coul = couleur du vin en question; $t_{stat} = 2.504$

ha = surface du vignoble; $t_{stat} = 2.743$

hacoul = surface du vignoble de la couleur du vin en question; $t_{stat} = -2.105$

age = âge moyen des vignes; $t_{stat} = 2.322$

vendmanu = vendange manuelle; $t_{stat} = 5.014$

prod = production moyenne; $t_{stat} = 2.083$

cotgm = cotation GaultMillau; $t_{stat} = 11.504$

cru = cru; $t_{stat} = 13.819$

t statistique de la constante $t_{stat} = -11.729$

$$R^2 = 0.57729$$

$$R_c^2 = 0.57116$$

$$F = 94.2312$$

$$DW = 1.215$$

écart-type de la régression = 11.486

prix moyen d'une bouteille de vin = 22.93 euros

Analyse des variables

Sur les neuf variables, on peut considérer qu'il y en six qui relèvent des coûts:

- 1) vigneron indépendant



de On pourrait se demander si "vigneron indépendant" est un critère
clair qualité ou de coûts, mais comme le coefficient de corrélation entre
cette variable et la cotation GaultMillau est de -0.08, il semble
que c'est un critère de coûts.

2) couleur du vin

des Du fait de la macération, il est logique q'un vin rouge engendre
coûts plus grands qu'un vin blanc.

3) surface du vignoble

4) surface du vignoble de la couleur du vin en question

variable La surface du vignoble a une influence positive, alors que la
première. suivante une influence négative qui pratiquement annule la

très Il en résulte que ces deux variables ensembles ont une influence
petite.

5) vendange manuelle

Là de nouveau la question de la qualité se pose, mais le coefficient
de corrélation avec la cotgm étant de 0.12, nous optons pour un
critère de coûts.

6) production moyenne

Dans une structure de coûts normale, cette variable devrait avoir
une influence négative, ce qui n'est pas le cas. Cette anomalie est
peut-être expliquée par la dispersion des parcelles

Trois qui relèvent de la qualité:

- 1) âge moyen des vignes
- 2) cotation GaultMillau
- 3) cru

Conclusion

On constate que ce sont surtout les critères de qualité qui influence le prix.
Par exemple, un grand cru est en moyenne environ 11 euros plus cher q'un premier
cru.

On pourrait être tenté d'essayer de deviner la cotation GaultMillau à travers
cette formule, mais statistiquement cette équation inverse n'est pas significative.

Cette communication sera étoffée dans les semaines à venir.