Contents

1 Introduction

2 Modélisation

2.1 Spécification du modèle

2.1.1 Choix des variables

2.1.2 Spécification du modèle

Le premier modèle est formulé de la sorte :

$$\begin{aligned} locataire_i = \beta_1 &+ \beta_2 diplomesup_i + \beta_3 jeunes_i + \beta_4 persagees_i \\ &+ \beta_5 appartements_i + \beta_6 chomage_i + \beta_7 urbanisation_i \\ &+ \beta_8 pauvrete_i + \varepsilon_i \end{aligned} \tag{1}$$

2.2 Première Estimation du modèle

2.2.1 Résultats et interprétation

On estime par la méthode des moindres carrés ordinaires les coefficients de l'équation précédente.

	Taux locataire	
Variable	Coefficient	Ecart-Type
Constante	-10.84531	13.04414
part_diplome_sup	0,4028115	0,0844738
part_jeune	0,9092502	0,285378
part_soixante_cinq_plus	0,1280139	0,2338338
part_appartement	0,1638682	0,0307458
taux_chomage	0,1978747	0,2883803
taux_urba	0,0080034	0,028728
taux_pauvrete	0,7533171	0,1317305
Nb. observations	96	
SCE	$3873,9635$	
SCR	$485,248588$	
SCT	$4359,20494$	

Le modèle s'écrit alors :

$$\begin{aligned} \hat{locataire}_i = -10.845 &+ 0.403 \times diplomesup_i \\ &+ 0.909 \times jeunes_i \\ &+ 0.128 \times persagees_i \\ &+ 0.164 \times appartements_i \\ &+ 0.198 \times chomage_i \\ &+ 0.008 \times urbanisation_i \\ &+ 0.753 \times pauvrete_i \end{aligned}$$

On peut calculer le coefficient de détermination R^2 :

$$R^2 = \frac{SCE}{SCT} = \frac{3873,9635}{4359,20494} = 0,8887$$

On calcule également le coefficient de détermination ajusté aux nombres de variables :

$$\bar{R}^2 = 1 - \frac{SCR/(N-K)}{SCT/(N-1)} = \frac{485,248588/88}{4359,20494/95} = 0,8798$$

Les deux coefficients sont élevés, cela laisse donc à supposer que l'ajustement de de la regression est de bonne qualité. Cependant, des tests de significativité (des paramètres et conjointe), et une étude de la multicolinéarité restent à être menés pour confirmer cela.

2.2.2 Tests de significativité

Dans un premier temps, on réalise un test F de significativité conjointe de tous les paramètres. L'hypothèse nulle, et l'hypothèse alternative de ce test sont telles que :

$$H_0 : \beta_2 = \beta_3 = \ldots \beta_K = 0$$
$$H_1 : \text{Au moins un } \beta_j \neq 0.$$

La statistique de test sous l'hypothèse nulle est distribuée selon une loi F de Fisher :

$$F = \frac{SCE/(K-1)}{SCR/(N-K)} \sim F(K-1, N-K)$$

On choisi un niveau de test à $\alpha = 5\%$. Et l'on compare la statistique calculée au quantile à 95% de la distribution F de Fisher avec comme degrés de liberte 7 et 88 respectivement au numérateur et au dénominateur.

$$F_{1-\alpha}(K-1, N-K) = F_{0.95}(7, 88) = 2.121$$

Après calculs, $F = 100,364$. La statisitique est supérieure au seuil, l'hypothèse nulle est rejetée au moins une variable permet d'expliquer le modèle.

On cherche maintenant à déterminer plus précisément quels sont les paramètres estimés significativement différents de 0. Pour cela, un test t de significativité est effectué sur sur les 8 paramètres :

$$H_0 : \beta_j = 0$$
$$H_1 : \beta_j \neq 0$$

La statistique de test calculée sous l'hypothèse nulle est distribuée selon une loi t de Student :

$$t_{\beta_j} = \frac{\hat{\beta}_j}{s_{\hat{\beta}_j}} \sim t(N-K)$$

On se place à un niveau de test bilatéral de $\alpha = 5\%$ et la statistique en valeur absolue doit être comparée au quantile à $97,5\%$ de la distributi on t de Student à 88 degrés de liberté, soit le seuil critique :

$$t_{1-\alpha/2}(N-K) = t_{0,975}(88) = 1,987289865$$

Tous calculs faits, on obtient les résultats suivants :

	Statistique t
constante	-.8314317
part_diplome_sup	4.768476
part_jeune	3.186125
part_soixante_cinq_plus	.547457
part_appartement	5.329783
taux_chomage	-.6861589
taux_urba	.2785922
taux_pauvrete	5.718624

Les statistiques de test $t_{\beta_1}, t_{\beta_7}, t_{\beta_6}$ et t_{β_3} sont inférieures en valeur en absolue au seuil. L'hypothèse nulle, est acceptée pour ces paramètres, ils ne sont pas significatifs. Les variables *urbanisation*, *chomage* et *persagees* seront donc retirées.
En revanche, $t_{\beta_2}, t_{\beta_3}, t_{\beta_5}, t_{\beta_8}$ sont supérieures en valeur absolue au seuil critique de $1,99$. On rejette l'hypothèse nulle pour ces paramètres ils sont significatifs, les variables *diplomesup*, *jeunes*, *appartements* et *pauvrete* permettent d'expliquer en partie *locataire*

2.2.3 Etude de la multicolinéarité

Une corrélation forte entre plusieurs variables explicatives peut induire une présence de multicolinéarité dans le modèle. Un calcul du VIF est fait pour chacune des variables.

Table 1: Facteur d'inflation de la variance

Variable	VIF
constante	0
part_diplome_sup	3,199862
part_jeune	12,48724
part_soixante_cinq_plus	14,58855
part_appartement	4,848903
taux_chomage	3,094149
taux_urba	4,416773
taux_pauvrete	2,663816

Il y a donc en effet de la multicolinéarité dans le modèle, les variables part_jeune et part_soixante_cinq_plus ont un VIF très élevé. Cette dernière sera donc retirée du modèle.

2.3 Estimation du modèle final

Un second modèle est formulé à la suite des résultats précédents. Ce modèle est tel que :

$$\begin{aligned} locataire_i = \beta_1 + \beta_2 diplomesup_i + \beta_3 jeunes_i \\ + \beta_4 appartements_i + \beta_5 pauvrete_i + \varepsilon_i \end{aligned} \tag{2}$$

2.3.1 Interprétation des résultats

2.3.2 Tests de significativité

2.3.3 Etude de la normalité des résidus

2.3.4 Test sur l'hétéroscédasticité

2.3.5 Estimation du modèle avec écart-type de White

3 Conclusion

dico : VIF = Variance inflation factor