Contents

1 Introduction

2 Modélisation

2.1 Spécification du modèle

2.1.1 Choix des variables

2.1.2 Spécification du modèle

Notre 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 l'équation. On obtient les résultats suivants :

Variables	Coefficients estimés	Écart-type
diplomesup	0.4028115	0.0844738
jeunes	0.9092502	0.285378
persagees	0.1280139	0.2338338
appartements	0.1638682	0.0307458
chomage	−0.1978747	0.2883803
urbanisation	0.0080034	0.028728
pauvrete	0.7533171	0.1317305
constante	−10.84531	13.04414

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

Avec :

$N = 96$ $K = 8$ $\quad SCE = 3873,9635 \quad SCR = 485,248588 \quad SCT = 4359,20494$

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 :

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

Le coefficient de détermination est élevé $R^2 = 0.89$, cela laisse penser que notre la combinaison linéaires de nos variables expliquent bien la part de locataires par départements. On effectue donc les tests de significativité des paramètres et de significativité conjointe pour prouver cela.

2.2.2 Tests de significativité

Dans un premier temps, nous allons effectuer un est F de significativité conjointe de tous les paramètres de la pente :

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

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 respectivement 7 et 88 degrés de liberté au numérateur et au dénominateur.

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

On cherche maintenant à savoir si les paramètres estimés sont significativement différents de 0. Pour cela on effectue le test suivant sur les $j = 8$ variables :

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

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 distribution t de Student à $N - K = 96 - 8 = 88$ degrés de liberté, soit le seuil critique :

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

Règle de décision :

$\text{Si}\lvert t_{\beta_j}\rvert < 1.99$	On accepte H_0
$\text{Si}\lvert t_{\beta_j}\rvert \geq 1.99$	On rejette H_0

Tous calculs faits, on obtient les résultats suivants :

Variables	t calculé
constante	$t_{\beta_1} = -0.83$
diplomesup	$t_{\beta_2} = 4.77$
jeunes	$t_{\beta_3} = 3.19$
persagees	$t_{\beta_4} = 0.55$
appartements	$t_{\beta_5} = 5.33$
chomage	$t_{\beta_6} = -0.69$
urbanisation	$t_{\beta_7} = 0.28$
pauvrete	$t_{\beta_8} = 5.72$

Les statistiques de test $t_{\beta_1}, t_{\beta_7}, t_{\beta_6} ett_{\beta_3}$ sont inférieures en valeur en absolue au seuil. On accepte donc l'hypothèse nulle, les paramètres ne sont pas significatifs, on va donc retirer les variables *urbanisation*, *chomage* et *persagees*.
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 (différents de zéro), les variables *diplomesup*, *jeunes*, *appartements* et *pauvrete* permettent d'expliquer en partie *locataire*

2.2.3 Etude de la multicolinéarité

2.3 Estimation du modèle final

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