Modèles de Régression

Projet HMMA307

Selena Iskounen

Sommaire 1

1. Comparaison modèle de régression à effet mixte et un modèle de régression linéaire
2. Comparaison de deux groupes à l'aide de la régression linéaire
3. Application de la causalité au sens de Granger

Modèle de régression à effet mixte VS Régression linéaire

Nous allons comparé le modèle de régression à un effet aléatoire au modèle de régression linéaire sur les données *aids* :

1. *id* : identité du patient,
2. *time* : temps de mort ou de censure,
3. *death* : décès pendant l'étude,
4. *obtime* : nombre de mois à partir de la première observation.
5. *drug* : type de traitement,
6. *sexe* : femme/homme,
7. *prevOI* : infection antérieur ou pas,
8. *AZT* : intolérance ou echec à l'AZT (zidovudine).

Modèle de régression à effet mixte VS Régression linéaire

Equation du modèle de régression à effet aléatoire :

$$y_{ij} = \beta_0 + \beta_1 \times x_{ij} + \beta_2 \times z_{ij} +_3 \times u_{ij} + \beta_4 \times s_{ij} + \beta_5 \times r_{ij} + \alpha_{ij} + \epsilon_{ij}$$

- β_o représente l'intercepte du modèle
- x_{ij} la variable *ostime*, z_{ij} la variable *drug*
- u_{ij} la variable *gender*, s_{ij} la variable *prevOI*
- r_{ij} la variable *AZT*, ϵ_{ij} les résidus
- α_{ij} l'effet aléatoire.

Modèle de régression à effet mixte VS Régression linéaire

Equation du modèle de régression linéaire :

$$y_{ij} = \beta_0 + \beta_1 \times x_{ij} + \beta_2 \times z_{ij} +_3 \times u_{ij} + \beta_4 \times s_{ij} + \beta_5 \times r_{ij} + \epsilon_{ij}$$

- β_o représente l'intercepte du modèle
- x_{ij} la variable *ostime*, z_{ij} la variable *drug*
- u_{ij} la variable *gender*, s_{ij} la variable *prevOI*
- r_{ij} la variable *AZT*, ϵ_{ij} les résidus
- α_{ij} l'effet aléatoire.

Modèle de régression à effet mixte VS Régression linéaire

Résultat du la régression a effet mixte sur les données aids

	Coef	Std.Err	z	P>|z|
Intercept	5.499	0.710	7.742	0.000
drug[T.ddI]	0.448	0.380	1.180	0.238
gender[T.male]	-0..306	0.652	-0.469	0.639
prevOI[T.noAIDS]	4.66	0.478	9.663	0.000
AZT[T.intolerantce]	0.261	0.472	0.554	0.579
Group var	15.253	0.683		

Modèle de régression à effet mixte VS Régression linéaire

Résultat de la régression linéaire sur aids

	Coef	Std.Err	z	P>|z|
obstime	-0.0840	0.025	-3.325	0.001
ddI	2.2923	0.135	16.937	0.000
male	1.6945	0.176	9.611	0.000
noAIDS	4.2548	0.176	24.217	0.000
intolerantce	2.1173	0.144	14.729	0.000

Modèle de régression à effet mixte VS Régression linéaire

Conclusion

Le modèle linéaire avec un effet mixte ne faisait pas mieux que le modèle de régression linéaire classique.

Comparaison de deux groupes grâce a une régression linéaire

Nous souhaitons savoir si il a une différence significatives entre les effets des variables suivantes sur le prix des logements de plus de 6 chambres (groupe B) et de moins de 6 chambres (groupe A) :

1. *CRIM* : variable correspondante au nombres de crimes.
2. *TAX* : variable correspondante au montant des taxes.
3. *LSTAT* : variable correspondante au statut sociale de propriétaire.

Comparaison de deux groupes grâce a une régression linéaire

Résultats de la régression linéaire :

	Coef groupe B	Coef groupe A
Intercept	37.5979	25.5981
CRIM	-0.02872	-0.1209
TAX B	0.000102	-0.000255
LSTAT B	-1.2238	-0.3719

Comparaison de deux groupes grâce à une régression linéaire

Conclusion

- Le nombre de crimes à un léger effet sur le prix des appartements du groupe B.
- Le montant des taxes à le même effet sur le prix des appartements des deux groupes.
- Le statut social du propriétaire à plus d'effet sur les appartements du groupe B.

Causalité au sens de Granger

Sur les données *ChickEgg* nous souhaitons savoir si le nombre d'oeufs engendre le nombre de poules et inversement. Pour répondre à cette problématique, nous allons utilisé le test de causalité au sens de Granger.

Les hypothèses :

H_o : les décalages du nombre d'oeufs n'engendrent pas des décalages sur le nombre de poulets.
H_1 : les décalages du nombre d'oeufs engendrent des décalages sur le nombre de poulets.

Causalité au sens de Granger

Résultat du test sur les données ChickEgg :

	Chicken	Egg
F	5.4050	0.5916
p-value	0.003	0.06238
df denom	44	44
df num	3	3

Causalité au sens de Granger

Conclusion

- Rejet de H_0 pour les poulets. (p-value < 0.05)
- Non rejet de H_0 pour les oeufs. (p-value > 0.05)

Nous ne pouvons pas affirmer qu'il y ait un effet de causalité entre les poules et les oeufs.

Résumé

Dans ce projet nous avons appliquer des modèles de régression pour répondre à des problématique tel que :

1. Quel est le plus performant entre un modèle de régression à effet mixte et un modèle linaire ?
2. Quels sont les variables qui ont un effet significatif sur plusieurs groupes ?
3. Quelle variable engendre l’autre sur une période donnée ?

Merci de votre attention !