

TP 2 - Régression logistique

Se dire montagnard : quels facteurs jouent sur ce sentiment ?

Une consultation en ligne *Montagne Debout* a été lancée en 2017 pour mieux comprendre le sentiment face à la montagne et la vision qu'ont d'elle les gens qui la pratiquent.

Une partie des réponses se trouvent dans la base de données *dataLogReg.csv*, avec les noms de variables suivants :

- *Montagnard* pour la réponse à la question : “Te sens tu montagnard ?” (Oui/Non)
- *Travailler* pour la réponse à la question : “Tu travailles en montagne” (0 = “Jamais”, 1 = “Une partie de l’année (saisonnier, ...)”, 2 = “Toute l’année”)
- *Habiter* pour la réponse à la question : “Habites tu en montagne ?” (Oui/Non)
- *Altitude* pour l’altitude du lieu de résidence estimé à partir du code postal des participants (en mètres).

1. Ecrivez la formule qui permet d’associer le fait de se sentir montagnard avec les variables *Travailler* et *Altitude*.
2. Appliquez le modèle proposé. Quelle est la taille de notre population ? Les variables sont-elles significatives d’après les sorties du modèle ?

Etudier l’association avec la variable *Travailler*

3. On regarde la variable catégorielle *Travailler*. Proposez un graphique de la distribution de cette variable.
4. Exprimez $OR_{1/0}$, $OR_{2/1}$ et $OR_{2/0}$ associés à la variable *Travailler* en fonction de $\hat{\beta}$ (par écrit). Qu’en déduisez vous de l’effet de cette variable sur le fait de se sentir montagnard ?
5. Recodez la variable *Travailler* en facteur :
Travailler = *factor(Travailler)*.

Appliquez ce nouveau modèle aux données. Les coefficients associés à la variable *Travailler* sont-ils significatifs ? Calculez leur intervalle de confiance avec la fonction *confint.default*.

6. Définissez une autre catégorie de référence pour la variable *Travailler* :
Travailler = *relevel(Travailler, ref="1")*.

Appliquez de nouveau le modèle (appelé *mod*). Les coefficients associés à cette variable sont-ils toujours significatifs ? Pourquoi et que pourriez-vous proposer ?

7. Utilisez la fonction suivante du package *aod* pour le test de Wald :
estBeta=*summary(mod)\$coefficients[,1]* # coefficients estimés
varBeta=*summary(mod)\$cov.unscaled* # matrice de variance covariance
wald.test(Sigma=varBeta,b=estBeta, Terms=4) # coefficient à tester

Obtient-on la même p-valeur que dans les sorties de la fonction *glm* ?

Appliquez un test de Wald pour la nullité simultanée de toutes les modalités de la variable *Travailler*. Cette variable est-elle significativement associée à notre variable expliquée ?

Etudier l'association avec la variable *Altitude*

8. On regarde maintenant la variable continue *Altitude*.
Proposez un graphique de la distribution de cette variable.
9. Exprimez $OR_{1000/500}$ et $OR_{500/0}$ associé à *Altitude* en fonction de $\hat{\beta}$ (par écrit). De façon plus générale que représente $OR_{1/0}$ pour *Altitude* ?
10. Pour relaxer cette hypothèse de linéarité de l'association, on pourrait passer la variable *Altitude* au logarithme. Cela permet-il d'améliorer la force de l'association entre cette variable et la variable expliquée ?

Etudier l'association avec la variable *Habiter*

11. On ajoute maintenant la variable *Habiter* au modèle. Ecrire la formule du modèle de régression logistique mis à jour.
Appliquez le modèle. Les coefficients associés à *Travailler* ont-ils changé ? Que peut-on dire de la variable *Habiter* par rapport à la variable *Travailler* ?
12. En posant $X_1 = \text{Travailler}$ et $X_3 = \text{Habiter}$, calculez $OR_{(X_1=1/X_1=0), X_3=0}$ et $OR_{(X_1=1/X_1=0), X_3=1}$ (i.e. OR associés à *Travailler* en fonction de la valeur de la variable *Habiter*). Qu'en concluez vous ?
13. On se propose d'ajouter un terme d'interaction $X_1 * X_3$ au modèle. Cela permet-il de résoudre le problème identifié à la question précédente ?
Appliquez ce nouveau modèle. Les termes d'interaction sont-ils significatifs ?

L'effet de faibles effectifs

14. Appliquer le modèle de la question 10 aux 100 premières observations du tableau de données. Que remarquez vous ?