

## 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 ?