

Remarques principales de la référente pédagogique (Myriam Vimond) :

- Lorsque le maître de stage n'est pas statisticien, il faut être régulièrement en contact avec la référente pédagogique et la contacter régulièrement lorsque l'on a un doute ;
- Lorsque l'on veut analyser les proportions le choix couramment utilisé pour la modélisation est une beta régression ;
- Dans le cas où l'on n'utilise pas cela il faut faire une revue de littérature qui permettra d'expliquer le choix de la spécification du modèle ;
- Le bilan personnel : est important de le faire mais ne doit pas être inclus dans le rapport ;
- Et pour finir, la remarque qui est la plus grave, j'ai commis une erreur lors de la modélisation du modèle de régression logistique (dans l'utilisation de la fonction glm, j'ai négligé un paramètre très important, ce qui fait que je n'ai pas utilisé la bonne fonction de lien).
  - ✓ Une première solution consistait à créer la fonction logit, calculer le  $\text{logit}(\text{PanS})$  avant de l'appliquer à mon cas (Je me rappelle avoir vu cela dans les ressources du JIACRA) ;
  - ✓ Une deuxième solution (la plus simple) était de tenir compte de ce paramètre, en mettant explicitement le nom de la famille dans glm (*family = binomial(link = "logit")*) et désactiver les avis (warnings).
- Faire une beta régression aurait permis de se rendre compte de cette erreur.

## Dans les populations animales

J'ai mis en œuvre une beta régression pour la famille animale avec une fonction de lien logit et j'ai obtenu les mêmes coefficients qu'une régression logistique. Cependant, la régression beta semble vraiment plus appropriée dans le sens où elle a une meilleure précision lorsque l'on analyse les P valeurs. Ici la validation du modèle de régression logistique se fait en analysant les déviations (si on regarde la P\_value, on aura tendance à conclure que l'association n'est pas significative).

En tenant compte de l'erreur (spécification du modèle) j'obtiens les coefficients ci-après pour les modèles de régression logistique ( $\text{logit}(\text{PanS}) = \alpha + \beta * \text{ALEA} + \varepsilon$ , où  $\varepsilon$  suit une loi normale de moyenne nulle et variance constante) dans les populations animales. Animale représente toutes les populations {et on a utilisé la spécification ayant pour effet fixe la population ( $\text{logit}(\text{PanS}) = \alpha + \beta * \text{ALEA} + \text{population} + \varepsilon$ ). La P valeur est celle du coefficient  $\beta$ .

Populations	Coefficient( $\beta$ )	Odds-ratio	P valeur
Animale	-0.62	0.53	0.71
Bovin	3.38	29.56	0.88
Chats et chiens	-4.46	0.011	0.787
Chevaux	-6.45	0.0015	0.787
Lapins	-0.37	0.69	0.879

Porcs	-0.46	0.62	0.903
Volailles	-0.92	0.39	0.75

Pour la régression beta, on obtient ceci :  $\text{logit}(\text{PanS}) = \alpha + \beta \cdot \text{ALEA} + \varepsilon$ ; où  $\varepsilon$  suit une loi beta de moyenne nulle et variance constante)

Populations	Coefficient ( $\beta$ )	Odds-ratio	P valeur
Animale	-0.62	0.53	1.80e-05 ***
Bovin	3.37	29.56	0.0164 *
Chats et chiens	-4.46	0.011	5.63e-05 ***
Chevaux	-6.43	0.0016	0.000253 ***
Lapins	-0.36	0.69	0.117
Porcs	-0.47	0.62	0.00046 ***
Volailles	-0.92	0.39	8.71e-07 ***

Dans les populations humaines :

La regression logistique :  $\text{logit}(\text{PanS}) = \alpha + \beta \cdot \log(\text{AMU}) + \text{annee} + \varepsilon$ , où  $\varepsilon$  suit une loi normale de moyenne nulle et variance constante.

Populations	Coefficient ( $\beta$ )	Odds-ratio
ES	-0.25	0.77
EHPAD	-0.33	0.71
Ville	-0.089	0.91

La régression beta avec fonction de lien logit également :

$\text{logit}(\text{PanS}) = \alpha + \beta \cdot \log(\text{AMU}) + \text{annee} + \varepsilon$ , où  $\varepsilon$  suit une loi beta.

Populations	Coefficients ( $\beta$ )	Odds-ratio	P valeur
ES	-0.25	0.77	2.21e-05 ***
EHPAD	-0.35	0.70	0.000232 ***
Ville	-0.089	0.91	1.33e-14 ***

Autres remarques :

1. On obtient les mêmes résultats en utilisant un modèle de régression logistique ou un modèle de régression beta avec la fonction de lien logit (dans notre cas)

2. Il existe plusieurs fonctions de lien pour la régression beta, et il faut choisir la meilleure spécification en y tenant compte.
3. Normalement, je devrais tenir compte des remarques pour la régression logistique en corrigeant mon rapport, mais je ne sais pas si cela serait encore utile vu que ces analyses seront refaites lors du projet statistique
4. En faisant la régression beta, il n'est plus utile de faire du Bootstrap dans le rapport pour avoir des intervalles de confiances précises.
5. Les codes R pourraient être améliorées en créant des fonctions (pour certaines tâches répétitives lors de la modélisation, vu qu'on fait pratiquement les mêmes analyses dans chaque population)

NB : le regression beta a été faite avec la fonction `betareg` de R. Les ressources contenues ici (permettent d'analyser et choisir la fonction de lien permettant d'avoir une meilleure précision) : <https://www.jstatsoft.org/article/view/v034i02>

(Cribari-Neto, F., & Zeileis, A. (2010). Beta Regression in R. *Journal of Statistical Software*, 34(2), 1–24. <https://doi.org/10.18637/jss.v034.i02>)