

Seeds: Random effect logistic regression

Lacoste-Badie Romane, Touzé Margaux & Loisel Camille

• Présentation des données

Pour ce projet, les données proviennent de la Table 3 de Crowder (1978). Elles concernent la proportion de graines ayant germé sur chacune des parcelles étudiées, réparties selon le type de graine et de racine. En effet, nous nous intéressons à des graines Bean et Cucumber pour des racines de type *O. aegyptiaco* 75 et *O. aegyptiaco* 73. Les valeurs sont les suivantes :

seed <i>O. aegyptiaco</i> 75						seed <i>O. aegyptiaco</i> 73					
Bean			Cucumber			Bean			Cucumber		
r	n	r/n	r	n	r/n	r	n	r/n	r	n	r/n
10	39	0.26	5	6	0.83	8	16	0.50	3	12	0.25
23	62	0.37	53	74	0.72	10	30	0.33	22	41	0.54
23	81	0.28	55	72	0.76	8	28	0.29	15	30	0.50
26	51	0.51	32	51	0.63	23	45	0.51	32	51	0.63
17	39	0.44	46	79	0.58	0	4	0.00	3	7	0.43
			10	13	0.77						

Figure 1 : Données Seeds

où r_i et n_i correspondent respectivement au nombre de graines qui ont germé et au nombre total de graines plantées sur la parcelle i .

Notons p_i la probabilité qu'une graine germe sur la parcelle i . On suppose ensuite :

$$r_i \sim \text{Binomial}(p_i, n_i)$$

$$\text{logit}(p_i) = \alpha_0 + \alpha_1 \cdot x_{1i} + \alpha_2 \cdot x_{2i} + \alpha_{12} \cdot x_{1i} \cdot x_{2i} + b_i$$

$$b_i \sim \mathcal{N}(0, \tau)$$

où x_{1i} et x_{2i} correspondent au type de graine et de racine de la parcelle i , avec un terme d'interaction $\alpha_{12} \cdot x_{1i} \cdot x_{2i}$. De plus, les paramètres $\alpha_0, \alpha_1, \alpha_2, \alpha_{12}$ et τ suivent des lois a priori indépendantes et non-informatives.

Ainsi, on obtient le modèle suivant :

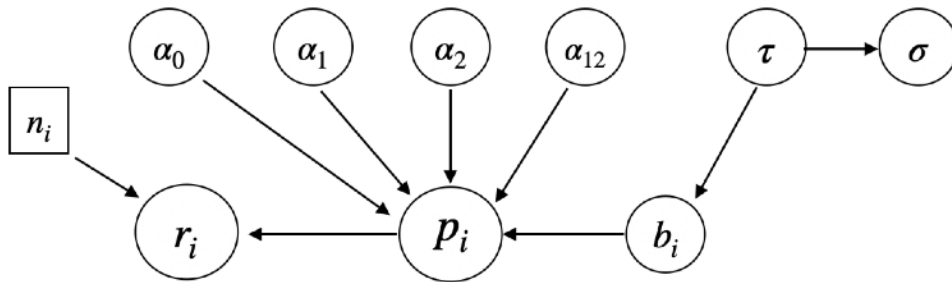


Figure 2 : Modèle Seeds

- Justification du modèle

Tout d'abord, le choix d'une loi binomiale pour r_i paraît naturel car il correspond au nombre de graines germées sur la parcelle i . Pour le calculer, on regarde bien le nombre de succès (la graine a germé) sur le nombre total n_i de graines plantées, sachant p_i , la probabilité qu'une graine germe.

Ensuite, pour p_i , on effectue une régression logistique avec les variables x_{1i} , x_{2i} , $x_{1i} \cdot x_{2i}$ et une constante afin de se ramener à des probabilités entre 0 et 1. On ajoute également un bruit b_i , propre à chaque parcelle i .

- Calculs mathématiques

L'objectif de cette partie est de calculer les lois conditionnelles de τ , α_0 , α_1 , α_2 , α_{12} et b_i .

Comme $\tau = \frac{1}{\sigma^2}$ où $\tau \sim \text{Gamma}(\alpha, \beta)$ avec $\alpha = \beta = 0.001$ et $b_i \sim \mathcal{N}(0, \tau)$, on a :

$$\begin{aligned} \pi(\tau | \alpha_0, \alpha_1, \alpha_2, \alpha_{12}, b_i, r_i) &\propto \pi(\tau) \cdot \pi(b_i | \alpha_0, \alpha_1, \alpha_2, \alpha_{12}, r_i, \tau) \\ &\propto \tau^{\alpha-1} \cdot \exp(-\beta\tau) \cdot \tau^{\frac{I}{2}} \cdot \exp\left(-\frac{1}{2}\tau \sum_{i=1}^I b_i^2\right) \\ &\propto \tau^{\alpha+\frac{I}{2}-1} \cdot \exp\left(-\tau \left(\beta + \frac{\sum_{i=1}^I b_i^2}{2}\right)\right) \end{aligned}, \text{ où } I = 21.$$

Ainsi, on obtient :

$$\tau \sim \text{Gamma}\left(\alpha + \frac{I}{2}; \beta + \frac{\sum_{i=1}^I b_i^2}{2}\right) \text{ avec } \alpha = \beta = 0.001.$$

De plus, comme $r_i \sim \text{Binomial}(p_i, n_i)$ et $\alpha_0 \sim \mathcal{N}(0, 10^6)$, on a :

$$\begin{aligned} \pi(\alpha_0 | \tau, \alpha_1, \alpha_2, \alpha_{12}, b_i, r_i) &\propto \pi(\alpha_0) \cdot \pi(r_i | \alpha_0, \alpha_1, \alpha_2, \alpha_{12}, b_i, \tau) \\ &\propto \exp\left(-\frac{\alpha_0^2}{2 \cdot 10^6}\right) \cdot \prod_{i=1}^I p_i^{r_i} \cdot (1 - p_i)^{n_i - r_i} \end{aligned}$$

Les paramètres α_1 , α_2 et α_{12} suivent la même loi normale que α_0 . Par conséquent, le calcul de leur loi conditionnelle est le même.

Enfin, pour b_i on obtient :

$$\begin{aligned} \pi(b_i | \alpha_0, \alpha_1, \alpha_2, \alpha_{12}, \tau, r_i) &\propto \pi(b_i | \tau) \cdot \pi(r_i | \alpha_0, \alpha_1, \alpha_2, \alpha_{12}, b_i, \tau) \\ &\propto \exp\left(-\frac{\tau}{2} \cdot b_i^2\right) \cdot \prod_{i=1}^I p_i^{r_i} \cdot (1 - p_i)^{n_i - r_i} \end{aligned}$$

- Analyse des résultats

	α_0	α_1	α_2	α_{12}	τ
Mean	-0.529584	0.070545	1.302763	-0.779451	92.451834
Sd	0.1503	0.2547	0.2432	0.3671	628.7747
Median	-0.538398	0.090572	1.304955	-0.794971	14.985754
Quantile 2.5%	-0.8070	-0.4479	0.7811	-1.4194	3.1845
Quantile 97.5%	-0.20764	0.54834	1.73289	0.06462	265.17082

Figure 3 : Estimation finale des paramètres

Les estimations des b_i sont visibles dans le Rmd joint à cette présentation. Ils sont globalement tous très proches de 0. Nous en concluons qu'il n'y a pas d'effet direct de la parcelle sur la probabilité de germination.

La constante α_0 vaut environ -0.53 . Le coefficient α_1 est très proche de 0 donc la variable x_{1i} , i.e le type de racine (*O. aegyptiaco* 75 et *O. aegyptiaco* 73), influence très peu la probabilité de germination. En revanche, le coefficient α_2 , associé au type de graine (Bean ou Cucumber), vaut 1.30. Le type Cucumber étant codé en 1, la probabilité de germination pour une graine Cucumber est plus élevée que pour une graine Bean. Enfin, le coefficient α_{12} vaut -0.78 , ce qui signifie que la probabilité de germination baisse lorsque la graine plantée est de type Cucumber et que la racine est de type *O. aegyptiaco* 73.

En résumé, nous recommandons de planter des graines Cucumber avec un type de racine *O. aegyptiaco* 75, pour avoir plus de chances de germination.