On s'intéresse à l'efficacité du vaccin contre la rougeole à partir de données épidémiques au Burundi (voir [1]). Trois classes d'âges sont considérées (première colonne du tableau 1) et on désigne par x_i le nombre de cas correspondant à la classe d'âge i parmi les N_{1i} enfants vaccinés et y_i le nombre de cas parmi les N_{0i} non vaccinés.

Classe d'âge (mois)	vaccinés		non vaccinés	
	x	N_1	y	N_0
Première classe : $[9-15]$	16	90	62	109
Deuxième classe : $[16-36]$	60	449	34	84
Troisième classe : $[37-60]$	33	413	3	19
Total	109	952	99	212

Table 1

La mesure standard de l'efficacité d'un vaccin s'obtient de la manière suivante : dans le cadre d'un modèle en *tout et rien* (appelé communément modèle II), il s'agit de la fraction π d'individus qui s'ils sont vaccinés, passent de *susceptible d'être infecté* à *complètement immunisé*. La fraction complémentaire $1-\pi$ a le même risque d'infection que la population non vaccinée. On désigne par p_1 la probabilité d'infection chez les vaccinés et p_0 la probabilité d'infection chez les non vaccinés (au cours d'une période donnée). Pour une classe d'âge i, on considérera donc p_{0i}, p_{1i} et π_i.

1. Préliminaires

(1) En supposant qu'il y a indépendance d'un individu à l'autre, quelle est la loi de probabilité du nombre de cas infectés chez les vaccinés et chez les non vaccinés dans une classe d'âge (c'est cette loi de probabilité qui sera utilisée dans la suite du devoir) ? Cette hypothèse vous semble-t-elle raisonnable pour les maladies infectieuses ?

(2) Écrire p_{1i} en fonction de p_{0i} et π_i et en déduire l'efficacité π_i en fonction de p_{0i} et p_{1i}. On notera cette fonction $h : \pi_i = h\big(p_{0i}, p_{1i}\big)$.

2. Homogénéïté par rapport à l'âge

On suppose que les probabilités d'infection p_{0i} et p_{1i} sont identiques et respectivement égales à p_0 et p_1 pour les trois classes d'âges.

(1) Écrire la vraisemblance $\ell(p_0, \pi)$ et la log-vraisemblance (on notera $x = x_1 + x_2 + x_3, y = y_1 + y_2 + y_3, N_1 = N_{11} + N_{12} + N_{13}, N_0 = N_{01} + N_{02} + N_{03}$).

(2) Dans un premier temps, on suppose que p_0 est une constante connue.

  (a) Calculer l'estimateur du maximum de vraisemblance $\widehat{\pi}_E$ de π en fonction de p_0.

  (b) Quelles sont l'espérance et la variance de l'estimateur $\widehat{\pi}_E$ en fonction de p_1 et p_0 puis de π et p_0 ?

  (c) Calculer l'information de Fisher $I(\pi)$ (en fonction de π et p_0). L'estimateur $\widehat{\pi}_E$ est-il de variance minimale ?

  (d) Quelle est la loi asymptotique de $\widehat{\pi}_E$?

  (e) On suppose que $p_0 = 50\%$. On se demande si ce vaccin a une efficacité supérieure à 60%.
  — Écrire les hypothèses nulle et alternative correspondantes.
  — Proposer un test basé sur la loi asymptotique de $\widehat{\pi}_E$.

— Effectuer le test de (en supposant les approximations asymptotiques légitimes) au risque 5%. Que concluez-vous ?
— Écrire la puissance de ce test pour $\pi = 75\%$ (On utilisera le résultat de la question (d)). Est-elle supérieure à 90% ?

(3) On suppose désormais que p_0 n'est pas connu. On veut calculer l'estimateur du maximum de vraisemblance $\widehat{\pi}_{\mathrm{MV}}$ de π. Cela sort du cadre du cours puisqu'il y a deux paramètres inconnus. Il s'agit donc de résoudre un système de deux équations (égalité à zéro des dérivées partielles par rapport à π et p_0) à deux inconnues. Procéder de la manière suivante :

(a) Écrire la dérivée de la log-vraisemblance obtenue à la question (1) par rapport à p_0 comme si π était une constante connue (il s'agit donc d'une dérivée partielle).

(b) Remplacer ensuite π par l'estimateur de maximum de vraisemblance $\widehat{\pi}_E$ fonction de p_0, obtenu à la question (2) (a).

(c) En déduire l'estimateur du maximum de vraisemblance $\widehat{p}_{0\mathrm{MV}}$ de p_0.

(d) En déduire l'estimateur du maximum de vraisemblance $\widehat{\pi}_{\mathrm{MV}}$ de π. On remarque que $\widehat{\pi}_{\mathrm{MV}}$ s'écrit en fonction de $\widehat{p}_{0\mathrm{MV}}$ et de $\widehat{p}_{1\mathrm{MV}} = \frac{x}{N_1}$. Quelle est sa valeur numérique ? Quelle est la valeur numérique du logarithme népérien de la vraisemblance maximale (aux constantes additives près) ?

3. Hétérogénéïté par rapport à l'âge

On considère que les probabilités d'infection peuvent être différentes suivant les classes d'âge, avec des efficacités spécifiques à chaque classe d'âge.

(1) Écrire la vraisemblance $L_i(p_{0i}, \pi_i)$ correspondant à la classe d'âge i.

(2) Écrire la log-vraisemblance $L_i(p_{0i}, \pi_i)$ et calculer l'estimateur du maximum de vraisemblance $\widehat{\pi}_{i\mathrm{MV}}$ de π_i.

(3) On se demande si on est dans la situation 2 (homogénéïté par rapport à l'âge) ou 3 (hétérogénéïté).

(4) Écrire les hypothèses nulle et alternative correspondantes.

(5) On souhaite tester ces hypothèses avec un test dit *de rapport de vraisemblances* (en supposant que l'on puisse utiliser les résultats asymptotiques) que nous allons construire ici.

(a) Écrire la vraisemblance $L(\pi_1, p_{01}, \pi_2, p_{02}, \pi_3, p_{03})$ sous H_1.

(b) Écrire le rapport des vraisemblances sous H_1 et H_0 [1] :

$$\frac{L(\pi_1, p_{01}, \pi_2, p_{02}, \pi_3, p_{03})}{L(p_0, \pi)}.$$

(c) En remplaçant p_{0i} et π_i par $\widehat{p}_{0i\mathrm{MV}}$ et $\pi_{i\mathrm{MV}}$, on trouve que le logarithme népérien de la vraisemblance maximale sous H_1, aux mêmes constantes additives près qu'à la question 3)(d) de la partie 2, vaut -473.21. Calculer la valeur numérique de la statistique du rapport de vraisemblances S donnée par

$$S = 2 \ln \left[\frac{L(\pi_1, p_{01}, \pi_2, p_{02}, \pi_3, p_{03})}{L(p_0, \pi)} \right],$$

en remplaçant chaque paramètre par la valeur de son estimateur par maximum de vraisemblance. [2]

1. Indication : La vraisemblance sous H_0 est calculée à la question 1) dans la partie 2.
2. penser à ce vous avez obtenu à la question 3)(d) de la partie 2

Nous avons ici 6 paramètres pour H_1 et 2 paramètres pour H_0 (qui est emboité dans H_1). La statistique de test S est distribuée sous H_0 suivant une loi du χ^2 dont le nombre de degrés de liberté est la différence $6 - 2 = 4$. Conclure au risque 5% [3].

Références

[1] C. M. Hernández-Suárez and C. Castillo-Chavez. Urn models and vaccine efficacy estimation. *Statistics in Medicine*, 19(6) :827–835, 2000.

3. Faire appel à un logiciel (R par exemple) pour calculer le quantile correspondant.