

1 Bonjour Charles,

2 J'ai effectué du « nettoyage » dans les logiciels statistiques avec mises à jour du document
3 de 140 page. Ce faisant, je me permets de te déranger. Bien entendu, sans urgence
4 aucune.

5 En effet, si tu as le temps dans les semaines à venir, je voudrais t'entretenir d'un soucis
6 « conceptuel » relatif au critère d'indépendance des données (cf. p.ex. chi2 ainsi que Logit)
7 pour lequel je ne trouve pas de « positionnement correct ».

8 En résumé, le problème est le suivant :

- 9 - 1 patient = 2 membres ; chaque membre = 1 à 8 (Na,C0..C6) signes cliniques CEAP
10 (Na=non affecté, ..., C3=œdème, ..., C6=ulcère, etc.)
11 - Totaux : 326 patients ; 724 membres ; signes cliniques CEAP : 876 (tous) mais 724 (si on
12 retient le maximum par membre).

13 **1/Problème vis-à-vis du chi2** : table (sexe, CEAP) : 2x8 ; si on maintient tous les CEAP, sur
14 les lignes M et F, on ajoutera donc dans certains cas 2 à n signes pour 1 même patient.

15 Question : les données sont-elles encore « indépendantes » au sens chi2 : un même
16 patient peut être source de plusieurs CEAP sur la ligne correspondante du tableau.

17 **2/Problème vis-à-vis du Logit** : CEAP = Logit(âge, sexe, membre) on a un df de 724
18 lignes (membres) ; sur chaque ligne, on a les binaires de présence de la variable
19 dépendante CEAP grâce aux colonnes NA, C0..C6 ; les colonnes des variables
20 indépendantes sont : âge et sexe et membre ; ici nous avons le cas où le patient est
21 « dupliqué » sur 2 lignes vis-à-vis de âge et sexe. Question : cette duplication des données
22 âge et sexe ne pose-t-elle pas un problème d'indépendance des données.

23 **Solutions en 1/ :**

24 - **ma crainte est infondée** : il y a indépendance des données parce que : p.ex. savoir que 3
25 patients ont soit C3 soit C6 soit C3&C6 est important au niveau du cout des traitements :
26 en retenant tous les CEAP, on peut déterminer si de manière statistiquement significative
27 (pval du chi2) les 2 C3 et 2 C6 le sont, peu importe que qu'1 C3 et qu'1 C6 proviennent d'un
28 même patient. Je compare cela à une chaine de production qui produit deux types de
29 pièces pouvant présenter chacune 1 à 8 défaut. Du point de vue des budgets de réparation,
30 je pense qu'il est exact de dresser un chi2 avec tous les défauts par pièce : la
31 comptabilisation de « l'œdème » (non masqué par « l'ulcère » sur la même pièce) me
32 semble une question statistiquement valable et non sujette à questionnement sur la
33 dépendance des données par rapport au même membre. Toutefois, je ne suis pas une
34 référence en statistiques et souhaite ton avis.

35 - **soit ma crainte est fondée** et il y a dépendance ; il faut se limiter à la solution CEAP Max

36 mais on perd de l'information (cf ci-dessus)
37 - **soit il existe des techniques** qui permettent de prendre en compte ce problème
38 - **soit chacune des approches** « CEAP tous » ou « CEAP Max » **est valable** (chi2 pval) et
39 peut être mise en œuvre parce que chacune répond en réalité à deux questions différentes.
40
41 **Solutions en 2/ :**
42 Même soucis mais cette fois entre membre et patient avec patient dupliqué
43 Avec tous mes remerciements.
44 Amitiés à tous,
45 Marc
46 PS :
47 1/ j'ai déjà fait des recherches mais les réponses varient du tout au tout.
48 2/ en dermatologie, les médecins « réduisent » le CEAP à 1 par patient : il méconnaissent le
49 membre à traiter et ignorent le cas de 2 membres simultanément malades
50 3/ en vasculaire, généralement ils prennent l'approche « CEAP max » par membre
51 4/ toutefois, et surtout du point de vue du Logit, ma curiosité scientifique me pousse à
52 questionner même l'approche 3/ : font-ils cela par facilité, par habitude ; surtout dans le
53 cas du Logit, leur position me paraît « intenable »