Bonjour Charles,

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

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

En résumé, le problème est le suivant :

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

**1/Problème vis-à-vis du chi2** : table (sexe, CEAP) : 2x8 ; si on maintient tous les CEAP, sur les lignes M et F, on ajoutera donc dans certains cas 2 à n signes pour 1 même patient. Question : les données sont-elles encore « indépendantes » au sens chi2 : un même patient peut être source de plusieurs CEAP sur la ligne correspondante du tableau.

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

**Solutions en 1/**:  
**- ma crainte est infondée** : il y a indépendance des données parce que : p.ex. savoir que 3 patients ont soit C3 soit C6 soit C3&C6 est important au niveau du cout des traitements : en retenant tous les CEAP, on peut déterminer si de manière statistiquement significative (pval du chi2) les 2 C3 et 2 C6 le sont, peu importe que qu’1 C3 et qu’1 C6 proviennent d’un même patient. Je compare cela à une chaine de production qui produit deux types de pièces pouvant présenter chacune 1 à 8 défaut. Du point de vue des budgets de réparation, je pense qu’il est exact de dresser un chi2 avec tous les défauts par pièce : la comptabilisation de « l’œdème » (non masqué par « l’ulcère » sur la même pièce) me semble une question statistiquement valable et non sujette à questionnement sur la dépendance des données par rapport au même membre. Toutefois, je ne suis pas une référence en statistiques et souhaite ton avis.  
**- soit ma crainte est fondée** et il y a dépendance ; il faut se limiter à la solution CEAP Max mais on perd de l’information (cf ci-dessus)  
**- soit il existe des techniques** qui permettent de prendre en compte ce problème  
**- soit chacune des approches** « CEAP tous » ou « CEAP Max » **est valable** (chi2 pval) et peut être mise en œuvre parce que chacune répond en réalité à deux questions différentes.   
  
**Solutions en 2/ :**  
Même soucis mais cette fois entre membre et patient avec patient dupliqué

Avec tous mes remerciements.

Amitiés à tous,  
Marc

PS :   
1/ j’ai déjà fait des recherches mais les réponses varient du tout au tout.  
2/ en dermatologie, les médecins « réduisent » le CEAP à 1 par patient : il méconnaissent le membre à traiter et ignorent le cas de 2 membres simultanément malades  
3/ en vasculaire, généralement ils prennent l’approche « CEAP max » par membre  
4/ toutefois, et surtout du point de vue du Logit, ma curiosité scientifique me pousse à questionner même l’approche 3/ : font-ils cela par facilité, par habitude ; surtout dans le cas du Logit, leur position me paraît « intenable »