Pourquoi la taille de l'échantillon est absolument indépendante de la présence de biais d'échantillonnage ?

Avant de commencer, une remarque préliminaire. Ecrire « l'échantillon est représentatif / n'est pas représentatif de la population » est faux, car ne voulant *rien* dire. En effet, supposons que la population cible d'une étude soit toutes les femmes adultes de France. Supposons maintenant un échantillon constitué d'étudiantes vétérinaires dans un amphi, que l'on va interroger par questionnaire. On pourrait se dire « mon échantillon des étudiantes vétérinaires n'est pas du tout représentatif de ma population des femmes adultes de France ». Et c'est faux. Pourquoi ? Imaginez que vous souhaitiez estimer la moyenne de la taille (en cm) des femmes adultes de France. L'estimation de la moyenne de la taille des étudiantes vétérinaires dans l'échantillon n'a *a priori* aucune raison d'être différente de celle de la population des femmes adultes de France ! Elle le sera à cause de la fluctuation d'échantillonnage, mais ce ne sera pas un écart systématique. En revanche, si vous souhaitez estimer la moyenne de l'âge des femmes adultes de France, là, évidemment, l'estimation dans l'échantillon va être systématiquement sous-estimée par rapport à celle de la population cible (les femmes adultes de France). Par conséquent, l'échantillon des étudiantes vétérinaires est *a priori* représentatif des femmes adultes de France en terme de taille, mais pas en terme d'âge. Donc, il faut *toujours* spécifier sur quel critère un échantillon est *a priori* représentatif d'une population, ou ne l'est *a priori* pas.

Ce qui conduit au fait qu'un échantillon n'est pas représentatif de la population cible sur un caractère particulier provient *exclusivement* d'un mauvais protocole d'échantillonnage. Reprenez l'exemple ci-dessous, pensez-vous que si on prend *toutes* les étudiantes vétérinaires des 4 écoles vétérinaires de France, l'échantillon va être davantage représentatif de la population des femmes adultes de France en terme d'âge ? Absolument pas !!! Dès lors qu'au départ, il y a mauvaise sélection, elle persistera indépendamment de la taille de l'échantillon !...

La seule situation où la taille de l'échantillon (de taille n) commence à être liée à la représentativité de la population sur un caractère particulier (de taille N), c'est lorsque $n/N > 0,1$. Mais en pratique, on n'a jamais une taille d'échantillon aussi grande, ou une taille de population aussi faible ! Il faut toujours partir du principe (qui est vrai) que la taille de l'échantillon sera *toujours* petite devant la taille de la population, quelle que soit la taille de l'échantillon que vous pourriez obtenir à partir d'une étude.