



L'épidémiologie pour tous

Fluctuations d'échantillonnage

Rédigée par B. Toma, B. Dufour, J. Rivière et J.J. Bénet ; validée par le bureau de l'AEEMA

Définitions

D'une manière générale, une **population** est « l'ensemble des êtres vivants présents dans un pays ou une région » et, au plan statistique « l'ensemble des unités soumises à une étude ».

L'épidémiologie a pour objet l'étude des maladies dans des populations.

Un **échantillon** est « un ensemble d'éléments sélectionnés à partir d'une population ». Un échantillon peut être **représentatif**, ce qui signifie qu'il correspond à un modèle réduit de la population de laquelle il est issu, ou non représentatif quand il ne « ressemble » pas à la population d'origine. A partir des résultats obtenus sur un échantillon représentatif, on peut estimer la valeur correspondante dans la population par inférence,

mais cette opération n'est pas opportune avec un échantillon non représentatif, en raison des biais possibles.

L'**échantillonnage** est « l'opération consistant à sélectionner des unités d'une population afin de former un échantillon ».

En épidémiologie, les études sont souvent effectuées sur des échantillons et, donc, on recourt fréquemment à l'échantillonnage.

Si par échantillonnage, on sélectionne plusieurs échantillons représentatifs à partir d'une même population, les résultats obtenus pour un même caractère dans chacun des échantillons sont fréquemment un peu différents entre eux et de la valeur réelle au sein de la population : ceci correspond aux **fluctuations d'échantillonnage**.



Les **fluctuations d'échantillonnage** sont définies comme « la **variation, liée au hasard, des écarts de composition, au regard du critère étudié, entre un échantillon et la population de laquelle il est issu** ».

Description des fluctuations d'échantillonnage

Si à partir d'une urne contenant 50 billes rouges et 50 billes blanches, on tire plusieurs fois un échantillon de 10 billes (avec remise dans l'urne après chaque

tirage), le nombre de billes rouges parmi les 10 tirées n'est pas toujours le même. Le plus souvent, il est de cinq, un peu moins souvent, il est de quatre ou de six, encore moins souvent, de trois ou de sept, etc.

Le nombre de billes rouges dans l'échantillon de 10 billes tirées au sort est d'autant moins fréquent (et, donc, moins probable) qu'il s'éloigne du nombre

correspondant à la composition de la « population » présente dans l'urne (50 billes rouges sur 100, soit une bille rouge sur deux). Ces écarts constatés entre les nombres de billes rouges trouvées dans différents échantillons de 10 billes tirées au sort, ainsi qu'avec le nombre de billes rouges correspondant à la composition de la population (50 % de billes rouges), illustrent les fluctuations d'échantillonnage. Ces écarts sont constatés quelles que soient la taille de l'échantillon et la composition d'une population hétérogène. Toutefois, l'amplitude des résultats obtenus sur différents échantillons de même taille est inversement proportionnelle à la taille des échantillons : ainsi, si les échantillons aléatoires sont d'un petit nombre d'unités, le résultat peut être uniquement des « billes rouges » ou uniquement des « billes blanches » ce qui laisserait penser à une population de 100 % de billes rouges ou de billes blanches. Si l'on continue de tirer des billes successivement, le nombre cumulé de billes rouges se rapproche progressivement de celui correspondant à la composition de la population pour atteindre, avec un nombre de billes tirées proche de la population totale, 50.

Origine des fluctuations d'échantillonnage

Les fluctuations sont dues au HASARD, le même qui conditionne la sortie de pile ou face pour une pièce de monnaie ou d'un chiffre de 1 à 6 pour un dé à jouer.

Il n'est pas possible de supprimer les fluctuations d'échantillonnage, pas davantage que de supprimer l'incertitude du résultat du lancer d'une pièce de monnaie ou d'un dé à jouer. En revanche, l'augmentation de la taille de l'échantillon (qui la rapproche de celle de la population) diminue l'amplitude de la gamme de résultats obtenus.

Conséquences de l'existence des fluctuations d'échantillonnage

Deux conséquences majeures de l'existence des fluctuations d'échantillonnage sont à prendre en considération en épidémiologie :

- Quelle que soit la qualité du tirage au sort d'un échantillon, le résultat obtenu sur cet échantillon, pour un caractère donné de la population, a un risque important d'être différent de celui de la population et, ce, d'autant plus que la taille de l'échantillon est réduite et faible par rapport à celle de la population. Il est donc indispensable de se rappeler que le résultat obtenu à partir d'un échantillon n'est qu'un élément servant à **estimer** le caractère dans la population, au sein d'une **fourchette** (calculée en fonction du risque accepté de se tromper par excès). Ce résultat ne peut pas, en effet, être considéré comme la valeur réelle du caractère dans la population à cause des fluctuations d'échantillonnage inhérentes au hasard, lié au tirage au sort des éléments de l'échantillon.
- Compte tenu de ces inévitables fluctuations d'échantillonnage, chaque fois qu'une **comparaison** est effectuée (notamment en épidémiologie descriptive et en épidémiologie analytique) entre des résultats obtenus à partir de deux (ou d'un plus grand nombre d') échantillons provenant de deux (ou d'un plus grand nombre de) populations, la question se pose de savoir si l'écart observé sur ces échantillons correspond à une réelle différence entre les populations d'origine ou si elle ne résulte que des fluctuations d'échantillonnage. Différents tests statistiques permettent d'apporter une réponse à cette question (avec un niveau d'erreur accepté).

En résumé : à cause des fluctuations d'échantillonnage, lors d'études sur des échantillons, il est nécessaire :



- D'estimer et d'**exprimer** le critère étudié sur un échantillon sous la forme d'une **fourchette de valeurs** appliquée à la population d'origine, avec un risque acceptable d'erreur par excès ;
- De soumettre les résultats obtenus sur deux (ou plus) échantillons contemporains de deux (ou plus) populations, ou successifs d'une même population, à un **test statistique** pour conclure, avec un risque acceptable d'erreur par excès, à l'existence, ou non, d'une **différence significative** entre les résultats.

Toma B. et al. : Epidémiologie appliquée, 2018, AEEMA éditeur, 614 p.

Terminologie : des liens hypertextes (en bleu souligné) vous permettent d'accéder aux définitions sur le site de l'AEEMA

Fiches AEEMA : ["p" ou degré de signification](#) ;

[Chi-deux](#).

