

comprendre l'épidémiologie

6. la différence est significative !

Bernard Toma

Service Maladies contagieuses
E.N.V.A.
7, avenue du Général de Gaulle
94704 Maisons-Alfort cedex

Objectifs pédagogiques

- Connaître le principe du χ^2 et son calcul.
- Savoir interpréter les résultats obtenus.

Définition

■ La table de χ^2 est un tableau de correspondance entre des probabilités d'obtention, et les valeurs d'un indice exprimant l'écart entre le résultat obtenu sur un échantillon et la valeur réelle.

Essentiel

■ Le principe du χ^2 consiste à déterminer un risque d'erreur accepté.

Cet article présente le principe du χ^2 , le calcul du χ^2 , les conditions de son application, l'interprétation de ses résultats.

Lors d'étude sur échantillons, il n'est pas possible d'emblée, à cause des fluctuations d'échantillonnage, d'affirmer qu'il existe une différence entre les résultats obtenus, pour une même caractéristique, au sein de deux populations.

Ainsi, si une enquête descriptive conduit aux résultats indiqués dans le **tableau 1**, comment être sûr que le taux d'élevages infectés dans le département A est plus élevé que dans le département B ? En effet, la "différence apparente" constatée (28 p. cent versus 24 p. cent) peut n'être due qu'au hasard, c'est-à-dire aux fluctuations d'échantillonnage qu'il entraîne.

● Avant de se prononcer, il va falloir comparer la différence apparente observée, avec les différences qui peuvent résulter des seules fluctuations d'échantillonnage.

● C'est le principe même d'un test statistique, en l'occurrence, le χ^2 (khi-deux ou khi-carré).

LE PRINCIPE DE χ^2

● Si, à partir d'une seule et même population, on estime la valeur d'une caractéristique de cette population (par exemple, le taux d'élevages infectés) sur deux échantillons représentatifs, la probabilité est élevée pour que, à cause du hasard, la valeur trouvée dans chacun des deux échantillons ne soit pas la même.

Exemple du tableau 1 :

Si dans le département A, une 2^e enquête était faite immédiatement sur un échantillon représentatif de 100 élevages, il est très probable que le nombre trouvé d'élevages à réponse positive soit différent de 28 ; il serait alors soit un peu plus élevé, soit un peu plus faible.

● Intuitivement, nous comprenons que la probabilité de trouver dans cette 2^e enquête un résultat proche de 28 est élevée, alors qu'un résultat éloigné de 28 est faible, et d'autant plus faible que plus éloigné de 28.

Tableau 1 - Résultats d'une enquête descriptive faite dans deux départements

Département	Élevages à réponse		Total
	positive	négative	
A	28	72	100
B	24	76	100
Total	52	148	200

● En recommençant ainsi de nombreuses fois, les résultats obtenus seraient distribués sous forme d'une courbe en cloche, centrée sur la valeur réelle dans la population du département A.

● À partir d'un modèle simple (par exemple, une urne contenant 50 p. cent de boules rouges et 50 p. cent de boules blanches), il est possible par tirages successifs d'échantillons, d'établir la distribution des résultats obtenus sur chaque échantillon et d'avoir la répartition des écarts (+1 ; +2 ; +3, ... ; -1 ; -2 ; -3, ...) par rapport à la valeur réelle dans la population (50 p. cent).

● Il est donc possible de déterminer les pourcentages d'obtention de tel ou tel écart à cause des seules fluctuations d'échantillonnage et, par conséquent, d'en déduire la probabilité d'obtenir telle ou telle valeur d'écart par rapport à la valeur réelle au sein de la population.

● Ceci permet d'aboutir à un tableau de correspondance (table de χ^2) (**définition**).

● Les probabilités sont élevées pour des écarts faibles entre le résultat obtenu sur un échantillon et la valeur réelle ; elles diminuent au fur et à mesure que les écarts augmentent.

● Le principe du χ^2 consiste à déterminer un risque d'erreur accepté (risque d'erreur par excès, c'est-à-dire d'affirmer l'existence d'une différence alors qu'il s'agit simplement de fluctuations d'échantillonnage).

Ce risque d'erreur correspond à une certaine valeur de l'indice exprimant l'écart. Il est souvent fixé à 5 p. cent.

- La valeur de l'indice découlant des résultats de l'enquête (comparaison des départements A et B) est ensuite comparée avec la valeur de l'indice correspondant au risque choisi.

FMC Vét