



# L'épidémiologie pour tous

## « $p$ » = degré de signification

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

Dans de nombreuses publications d'épidémiologie (en particulier d'épidémiologie analytique ou explicative), ainsi que pour d'autres domaines scientifiques, on trouve, à côté de résultats chiffrés, la lettre «  $p$  » accompagnée du signe «  $<$  » ou «  $=$  » suivi d'une valeur numérique.

Il s'agit du degré de signification (en anglais : *p value*, *probability value* ou *observed significance level*), notion considérée comme indispensable pour renseigner sur la fiabilité des informations apportées par la publication.

De quoi s'agit-il ?

## La problématique

### En épidémiologie analytique

En épidémiologie analytique, la démarche habituelle est d'étudier la validité de l'hypothèse de facteurs de risque supposés. La première opération consiste à vérifier l'existence d'une association entre un (des) facteur(s) de risque supposé(s) et un état pathologique. Cette vérification passe par la **comparaison** entre deux groupes de sujets pour voir si les malades et les non malades ont la même fréquence d'exposition au facteur de risque supposé (étude cas/témoins) ou si la maladie a ou non la même fréquence d'apparition selon que les sujets sont exposés ou non au facteur de risque supposé (étude exposés/non exposés).

Le premier objectif de l'étude est de déterminer s'il **existe ou non une différence entre les deux groupes comparés**. La difficulté pour conclure sur cette question est due au fait que les fluctuations d'échantillonnage entraînées par le hasard peuvent

induire des différences dans les résultats obtenus à partir d'échantillons provenant de la même population. Si l'écart numérique constaté entre les deux groupes comparés est considérable, la conclusion est facile. Mais, si cet écart n'est pas très grand, il existe un risque de prendre pour une différence ce qui n'est que le résultat du seul hasard sur l'échantillonnage. Il faut donc essayer de trouver une règle de décision permettant, en présence d'un écart constaté entre les deux groupes comparés, de considérer que les caractéristiques de ces deux groupes ne sont pas identiques (présence d'une différence) ou bien, au contraire, qu'il n'est pas possible d'être sûr de l'existence d'une différence.

### En épidémiologie descriptive

En épidémiologie descriptive, on peut avoir besoin de comparer deux situations (par exemple le pourcentage de prévalence d'une maladie dans deux régions, ou son évolution au cours du temps dans une même région), ou de chercher si des associations peuvent être mises en évidence entre la maladie et diverses caractéristiques des groupes d'individus, en vue d'une éventuelle étude analytique ultérieure. Le même type de question se pose : l'écart constaté correspond-il à une différence réelle ou bien est-il dû simplement à des fluctuations d'échantillonnage ?

## Approche statistique simplifiée

Soit une urne contenant 50 % de boules blanches et 50 % de boules rouges. Le tirage au sort d'un échantillon de 20 boules peut conduire à obtenir un nombre varié de boules rouges. Ce nombre est centré

sur la valeur de la proportion de boules rouges dans l'urne (50 %, donc 10 boules rouges dans l'échantillon) mais il peut être supérieur ou inférieur à 10, avec une fréquence d'autant plus faible que l'on s'éloigne de 10. Il est possible, par des tirages successifs, de déterminer la fréquence (et donc la probabilité) avec laquelle un résultat différent de 10 est obtenu, dû au hasard (fluctuation d'échantillonnage). On dispose alors, pour chaque nombre de billes rouges obtenues par tirage au sort d'un échantillon de 20 boules, de la probabilité de son occurrence sous la seule influence du hasard.

Si à partir d'une autre urne, contenant aussi des boules blanches et des boules rouges, mais en proportions inconnues, un tirage semblable est effectué, la question qui se pose, à la vue du nombre de boules rouges dans l'échantillon de 20 boules (par exemple, 7 sur 20), est : la proportion de boules rouges dans l'urne numéro 2 est-elle différente ou non de celle de l'urne 1 ? Soit deux hypothèses :  $H_0$ , pas de différence ;  $H_1$ , présence d'une différence.

Autrement dit, si, compte tenu du résultat (7 boules rouges sur 20), je **conclus à une différence** de proportion de boules rouges dans les deux urnes : quelle est la probabilité pour que le hasard soit responsable, à lui seul, de cet écart et que je fasse une **erreur « par excès »** (erreur de première espèce ou erreur alpha) en considérant que la proportion est différente ? Bien sûr, si je conclus, au contraire, que je ne peux pas rejeter l'hypothèse  $H_0$  et, donc, qu'il n'y a **pas de différence**, je peux faire une autre **erreur, « par défaut »** (erreur de deuxième espèce ou erreur bêta) si, en fait, il y a par exemple 60 % de boules blanches et 40 % de boules rouges dans l'urne 2.

## Le pari

Le recours à un test statistique ( $\chi^2$  par exemple) permet de répondre et de rejeter ou non  $H_0$  en fonction du résultat du test qui quantifie l'importance de l'écart numérique entre les deux groupes comparés (tableau 1).

**Tableau 1 : Représentation des différentes conclusions d'un test statistique en fonction de la réalité supposée connue**

		Réalité	
		Ecart important	Ecart négligeable
Conclusion retenue	Ecart important (> seuil)	<b>Exact</b>	<b>Erreur de 1<sup>ère</sup> espèce (alpha)</b>
	Ecart négligeable (≤ seuil)	<b>Erreur de 2<sup>ème</sup> espèce (bêta)</b>	<b>Exact</b>

Il est nécessaire de **fixer a priori le niveau de risque d'erreur par excès (risque d'erreur alpha) acceptable**, c'est-à-dire le risque de rejeter  $H_0$  et de conclure à tort à l'existence d'une différence (alors qu'elle n'existe pas) : il s'agit du « **seuil de signification** », le plus souvent fixé à 0,05. Autrement dit, j'accepte un risque de 5 % de conclure, en fonction de l'écart observé, qu'il existe une différence alors que cet écart est dû au seul hasard. Bien évidemment, la valeur du « seuil de signification » retenu peut être différente en fonction de l'objectif du chercheur : plus élevé (par exemple, 0,10), en

première approche, parce que le résultat dépend du nombre de sujets étudiés et que le nombre utilisé est faible ; plus faible (par exemple, 0,01), pour avoir une plus grande probabilité que le hasard n'est pas en jeu.

## Calcul de « p »

La valeur de « p », calculée à partir des observations (donc *a posteriori*) et exprimée en valeur décimale, est fournie par des tableaux et/ou par des logiciels (cf. site AEEMA : [Tableur pour réaliser un test du Chi 2](#)).

Elle peut varier entre 1 et 0.

## Interprétation de « p »

- La valeur de « p » indique la **probabilité** que dans la situation donnée, en l'absence de biais et d'erreur, **le seul hasard soit responsable de l'écart observé**.
- **Si « p » est inférieur au seuil de signification (risque alpha) fixé, la différence est considérée comme significative.**

Exemple :  $p = 0,02$  ; si le seuil de signification fixé *a priori* est de 0,05, on conclut à l'existence d'une différence significative au seuil de risque alpha de 0,05. Si le seuil fixé était de 0,01, la conclusion serait différente.

Plus « p » a une valeur faible, plus la probabilité (en l'absence d'erreur et de biais) que l'écart observé soit le fruit du seul hasard est faible. Cette valeur ne doit pas pour autant être considérée comme le risque de conclure à tort par excès, mais seulement comme la valeur permettant de juger si l'écart est ou non significatif par rapport au seuil de signification accepté, défini *a priori*.

- Lorsque « p » est supérieur au seuil de signification fixé, il n'est pas possible de rejeter l'hypothèse nulle et de considérer qu'il existe une différence. Bien sûr, une différence peut néanmoins exister mais être trop faible, ou le

nombre de sujets étudiés trop petit, pour qu'elle soit mise en évidence.

Exemple :  $p = 0,08$  ; si le seuil de signification fixé *a priori* est de 0,05, il n'est pas possible de rejeter  $H_0$  et donc de considérer qu'il existe une différence. Cependant, si le seuil de signification fixé *a priori* était de 0,10, il serait possible de rejeter  $H_0$ .

Le choix de la valeur seuil doit avoir été fait avant que les résultats aient été obtenus, pour éviter que ce ne soit cette valeur qui conduise à ce choix, enlevant alors son objectivité scientifique à l'observateur. En règle générale, une valeur supérieure à 0,05, comme 0,10 ou 0,20, est usuelle dans les démarches exploratoires, pour éviter de ne pas prendre en considération une association qui pourrait être intéressante pour la suite de l'analyse.

## Généralisation de l'usage de « p »

L'énorme intérêt de l'usage de la valeur « p » est de rendre l'interprétation du résultat indépendante du type de test statistique : quel que soit le test, aussi sophistiqué soit-il, l'usage reste le même, « p » permettant de situer le résultat du test par rapport à la valeur de risque alpha retenu (seuil de signification) et donc d'en déduire l'interprétation. C'est l'usage dans les communications scientifiques.

*Lors de comparaison, un seuil de signification du risque d'erreur alpha (erreur par excès) doit être fixé a priori.*



*Un degré de signification : « p » est calculé a posteriori à partir des résultats obtenus, que l'on compare avec le seuil de signification :*

- Si « p » < seuil de signification (risque d'erreur alpha) : **différence significative**
- Si « p » > ou = seuil de signification (risque d'erreur alpha) :  **$H_0$  non rejetée ; pas de différence constatée.**

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

Annexe : [Tableur pour réaliser un test du Chi 2](#)



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

Fiches AEEMA : [Fluctuations d'échantillonnage](#) ;  
[Chi-deux](#).