

**DEGRÉ DE SIGNIFICATION STATISTIQUE « P »
EN MÉDECINE**

**Oifa SAIDI (*), Nadia BEN MANSOUR, Said HAJEM,
Mohamed HSAIRI**

INTRODUCTION

Les chercheurs dans le domaine scientifique se posent souvent la question de savoir si la différence entre des paramètres estimés (proportions, taux, moyennes etc.) est statistiquement significative et, le cas échéant, quelle est la valeur du degré de signification (petit p).

Si l'on prend l'exemple d'un essai clinique randomisé auprès de 200 patients atteints d'ulcère duodéal, répartis en deux groupes égaux de 100 patients chacun, l'un recevant un traitement A, et l'autre un traitement B. Au bout de six semaines de traitement, les pourcentages de guérison sont de 45% pour le traitement A et 23% pour le traitement B.

Dans ce cas de figure, on est en droit de se demander si la différence observée entre les deux pourcentages de guérison est réelle ou due uniquement au hasard qui aurait pu favoriser le traitement A en lui allouant un plus grand nombre de patients répondeurs.

Tests d'hypothèses

Répondre à la question précédente revient à choisir entre deux hypothèses :

- ✓ Première hypothèse : Les deux proportions sont identiques ($P_A=P_B$) et les deux traitements A et B sont équivalents. Cette hypothèse est appelée hypothèse nulle, symbolisée par H_0 .

*) Statisticienne, Institut National de Santé Publique

- ✓ Deuxième hypothèse : Les deux proportions sont différentes ($P_A \neq P_B$) et les deux traitements A et B ne sont pas équivalents. Cette hypothèse est appelée hypothèse alternative, symbolisée par H_1 .

Intuitivement, plus l'écart observé entre les 2 pourcentages estimés est grand, plus faible est la probabilité que la différence observée résulte uniquement du hasard. Il convient toutefois de remarquer qu'un faible écart observé entre deux pourcentages estimés n'exclut pas qu'ils soient différents.

Le test statistique permet de formaliser ce raisonnement intuitif. Il aidera à trancher entre les deux hypothèses nulle (H_0) et alternative (H_1). En supposant au préalable que H_0 est vraie, le test statistique permet de quantifier la probabilité d'obtenir une différence égale ou supérieure à celle entre P_A et P_B . Si cette probabilité est faible, nous affirmons que la différence observée ne peut pas être attribuée au seul hasard et nous en concluons qu'elle est significative.

Les risques d'erreurs α et β

La conclusion qui devrait permettre de trancher entre les deux hypothèses nulle et alternative est toujours entachée d'erreurs. Il existe deux types d'erreurs notés α (erreur de première espèce) et β (erreur de deuxième espèce).

i) L'erreur de première espèce α est la probabilité de conclure à une différence entre les deux pourcentages (rejeter H_0 alors que H_0 est vraie). Le risque α est encore appelé seuil de signification du test. Ce dernier représente le seuil à partir duquel on considère que la différence observée lors d'un test est statistiquement significative.

Classiquement, le seuil de la signification statistique est fixé à 5%. Dans certaines circonstances, ce seuil peut être abaissé à 1%, si l'on estime qu'un niveau d'erreur de 5% est trop élevé eu égard à l'importance des conclusions à tirer des résultats. C'est le cas, par exemple, d'un test visant l'évaluation de l'efficacité d'un vaccin destiné à être très largement diffusé.

ii) L'erreur β de deuxième espèce est la probabilité de conclure à l'absence de différence entre les deux pourcentages (accepter H_0 alors que H_0 est fautive). La probabilité $1 - \beta$ est appelée puissance du test en ce sens qu'elle exprime la capacité du test à mettre en évidence une différence qui existe réellement.

Le calcul du degré de signification «p»

Le choix entre les deux hypothèses nulle et alternative se fait à travers le calcul d'un paramètre désigné par la lettre «p» (p-value selon les anglo-saxons). Ce paramètre désigne la probabilité d'observer des résultats au moins en désaccord avec l'hypothèse nulle que ceux qui ont été obtenus.

Les logiciels d'analyse statistique permettent de calculer aisément la valeur de cette grandeur. Dans le cas de l'exemple précédent concernant la comparaison des deux traitements A et B, le degré de signification p est inférieur à 0.001. Cette valeur est très largement inférieure au seuil de signification traditionnellement choisi (5% ou 1%). La différence observée est, par conséquent, hautement significative. Il est peu probable que cette différence soit due au seul hasard.

Si cette valeur p est supérieure à la valeur du seuil préalablement défini (habituellement 5% ou 1%), on affirme que la différence est non significative. La probabilité qu'elle soit due au seul hasard est forte.

De façon générale, plus le degré de signification est faible, plus on est convaincu que les résultats observés ne sont pas en cohérence avec l'hypothèse nulle et que la différence observée est bien réelle.

Il convient également de noter que le degré de signification nous permet d'affirmer avec plus ou moins de conviction qu'il y a une différence, mais en aucun cas il ne nous renseigne sur l'importance de cette différence. En effet, la valeur de p dépend de la différence observée entre les deux groupes et de la taille d'échantillon. S'il existe une différence réelle, aussi infime soit-elle entre 2 groupes, il est fort probable que n'importe quel test statistique aboutisse à une valeur de p inférieure à 0.05, dès lors que le nombre de sujets étudiés est important.

Notons enfin qu'un résultat non statistiquement significatif peut être imputable aux deux causes suivantes :

- ✓ l'hypothèse H_0 est vraie (i.e. il y a équivalence entre les deux groupes).
- ✓ la puissance statistique n'est pas suffisante (i.e. nombre de sujets insuffisant).

CONCLUSION

Le calcul du degré de signification « p » aide à trancher entre l'hypothèse nulle H_0 d'égalité des paramètres et l'hypothèse alternative H_1 stipulant l'existence d'une différence. La valeur de « p » devrait être interprétée avec précaution. Une valeur inférieure au seuil de signification préalablement défini (5% ou 1%) permet de conclure à l'existence d'une différence significative entre les paramètres à comparer. En revanche, une valeur supérieure à ce seuil ne permet pas de rejeter l'hypothèse nulle.

Il faut insister sur le fait que la valeur du degré de signification dépend de la puissance du test statistique. Sachant que celle-ci dépend de la taille des groupes, un nombre réduit de sujets inclus peut empêcher de détecter statistiquement une différence qui existe réellement.

Il convient enfin de noter que toute différence statistiquement significative n'est pas forcément cliniquement significative. La signification statistique n'implique donc pas forcément la pertinence clinique.

RÉFÉRENCES

- Bouyer J. Méthodes statistiques. Médecine- Biologie. Editions INSERM. Paris, 1996.
- Ancelle T. Statistique Epidémiologie. 3^eédition. Paris 2011.
- Armitage P, Berry G. Statistical methods in medical research. Blackwell Scientific Publications. Oxford 1971.
- Schwartz D. Méthodes en épidémiologie. Flammarion Médecine sciences. Paris 1970.
- Daures JP. Probabilités et statistiques en Médecine. Edition Sauramps médical. Montpellier 1993.

