

Qu'est ce que le « p value » ?

Marc Girondot

UNIVERSITÉ PARIS-SACLAY FACULTÉ DES SCIENCES D'ORSAY

1

Définitions verbales

- Dans un test de type H_0-H_1 , l'hypothèse H_0 est dite hypothèse nulle. Elle est clairement explicitée. L'hypothèse H_1 est l'hypothèse alternative. Ce sont toutes les hypothèses qui ne sont pas H_0 .
- Quand on fait un test, on cherche si le jeu de données a pu être obtenu sous l'hypothèse H_0 . Il faut noter que la réponse est tout le temps « oui » mais avec une probabilité plus ou moins élevée. Le « p value » désigne cette probabilité (Sir Ronald Fisher, *Statistical Methods for Research Workers*, 1925).

H_0 : Hypothèse nulle
 H_1 : Hypothèse alternative

x : jeu de données
 p : probabilité que le jeu de données x ait pu être obtenu avec H_0

2

Définitions verbales

- Donc le « p value » représente la probabilité de se tromper si on rejette H_0 . Par exemple, si $p=0,2$, cela signifie que si on rejette H_0 , on sait que ce jeu de données avait 20% de chance d'être obtenu alors que H_0 était vraie. Donc on prend un risque assez grand en rejetant H_0 . Il faut dire qu'on ne rejette pas H_0 .
- *Attention: on n'accepte jamais H_0 car on ne peut jamais rejeter H_1*

3

Les valeurs de p

Si un effet possède un niveau p de 0,019, et qu'un autre possède un niveau p de 0,048, il est incorrect de conclure que nous avons la preuve statistique que le premier effet est plus fort que le second.

Pour interpréter correctement ces résultats, nous avons besoin de plus d'informations.

Pour comprendre pourquoi, supposons qu'un article reporte une valeur p de 0,001. Ce niveau p peut être lié à un effet évident dans la population avec une taille d'échantillon importante, ou un effet important dans la population avec une taille d'échantillon modérée, ou encore avec un effet très important dans la population avec une petite taille d'échantillon.

De la même manière, un niveau p de 0,075 peut représenter un effet important combiné à une taille d'échantillon faible, ou un effet minuscule avec un échantillon de très grande taille.

Il faut donc prendre beaucoup de précautions pour comparer des niveaux p.

4

Définition

La p-value est donc, d'un point de vue formel: $\text{prob}(x|H_0)$, le signe | signifiant « sachant »;

C'est donc la probabilité d'observer les données x sachant que H_0 est vrai.

Mais est-ce vraiment ce qu'on veut savoir ?

5

On cherche en fait...

- $\text{prob}(x|H_0)$ ne nous donne pas la probabilité que H_0 soit vraie mais la probabilité que les données x aient pu avoir été obtenues sous H_0 .
- Or on est intéressé par $\text{prob}(H_0|x)$ ce qui se lit: quelle est la probabilité de H_0 sachant qu'on a observé les données x .

• Hubbard, R., Lindsay, R.M., 2008. Why P values are not a useful measure of evidence in statistical significance testing. *Theory & Psychology* 18, 69-88.

6

D'après les règles de bases de probabilités

$\text{prob}(H_0, x) = \text{prob}(H_0) \text{prob}(x|H_0)$

$\text{prob}(H_0, x) = \text{prob}(x) \text{prob}(H_0|x)$

Donc:
 $\text{prob}(H_0|x) = \text{prob}(H_0, x)/\text{prob}(x)$
 $\text{prob}(H_0|x) = \text{prob}(x|H_0) \cdot \text{prob}(H_0)/\text{prob}(x)$

Donc:
 $\text{prob}(H_0|x) \neq \text{prob}(x|H_0)$

7

Journal of the American Statistical Association

$\text{prob}(H_0|x)$ vs. $\text{prob}(x|H_0)$

- En utilisant un test statistique Bayésien sur une moyenne normale, James Berger et Thomas Sellke ont montré que pour des valeurs de p soit $\text{prob}(x|H_0)$ de 0,05, 0,01, et 0,001, les probabilités postérieures du modèle nul, $\text{prob}(H_0|x)$, pour $n = 50$ sont 0,52, 0,22, et 0,034. Pour $n = 100$ les valeurs étaient de 0,60, 0,27, et 0,045.

Berger, J.O., Sellke, T., 1987. Testing a point null hypothesis: The irreconcilability of p values and evidence (with comments). *Journal of the American Statistical Association* 82, 112-139.

8

Royal Society
Open Science

- If you use $p = 0.05$ to suggest that you have made a discovery, you will be wrong at least 30% of the time.

Colquhoun, D., 2014. An investigation of the false discovery rate and the misinterpretation of p-values. Royal Society Open Science 1, 140216-140216.

9

PNAS...

- Les progrès récents des tests d'hypothèse bayésiens ont conduit au développement de tests bayésiens qui ont les mêmes régions de rejet que les tests de signification classiques.
- Sur la base de la correspondance entre ces deux classes de tests, il est possible d'assimiler la taille des tests d'hypothèses classiques aux seuils de preuve dans les tests bayésiens, et d'assimiler les valeurs P aux facteurs de Bayes.
- Un examen de ces liens suggère que les préoccupations récentes concernant le manque de reproductibilité des études scientifiques peuvent être attribuées en grande partie à la conduite de tests de signification à des niveaux de signification injustifiés.
- Pour corriger ce problème, les seuils de preuve requis pour la déclaration d'un résultat significatif devraient être augmentés. En termes de tests d'hypothèses classiques, ces normes de preuves imposent la conduite de tests au niveau de signification de 0,005 ou 0,001.

Johnson, V.E., 2013. Revised standards for statistical evidence. Proceedings of the National Academy of Sciences of the United States of America 110, 19313-19317.

10

Nature...

- P values, the 'gold standard' of statistical validity, are not as reliable as many scientists assume.

Nuzzo RL (2014) Scientific method: Statistical errors. Nature 506:150-152

11

American
Statistician
Association

- Q: Why do so many colleges and grad schools teach $p = .05$?
- A: Because that's still what the scientific community and journal editors use.
- Q: Why do so many people still use $p = 0.05$?
- A: Because that's what they were taught in college or grad school.

Baker M (2016) Statisticians issue warning on P values. Nature 351:151-152
Wasserstein RL, Lazar NA (2016) The ASA's statement on p-values: context, process, and purpose. The American Statistician 70:129-133

12

Erreurs classiques et mauvaises interprétations sur les p-values

- La p-value n'est pas la probabilité que l'hypothèse nulle soit vraie, ni la probabilité que l'hypothèse alternative soit fausse
- La p-value n'est pas la probabilité que les données soient dues simplement à un coup de chance.
- La p-value n'est pas la probabilité de rejeter à tort l'hypothèse nulle.
- La p-value n'est pas la probabilité que reproduire l'expérience donnerait la même conclusion.
- Le seuil de signification, par exemple 0,05, n'est pas déterminé par la valeur de p.
- La p-value ne représente pas la force ou l'importance de l'effet observé.

13

Pourquoi maintenant ?

- Le problème lié à l'utilisation des p-values est connu depuis longtemps dans le domaine des statistiques mais n'a percolé que récemment en dehors de ce domaine.
- Pour comprendre ce phénomène, il faut relire Tomas Kuhn, La Structure des Révolutions Scientifiques. Il y démontre notamment comment un paradigme en vient à en remplacer un autre.

The Structure of Scientific Revolutions est un essai rédigé par le philosophe et historien des sciences, Thomas Samuel Kuhn (1962).

14

Alors, que faire ?

- *CONSTRUCTION DE MODÈLES BAYÉSIENS*
- *COMPARAISON DES INTERVALLES CRÉDIBLES*

15