

La fin de la valeur p binaire

Pier-Olivier Caron¹

¹Université TÉLUQ

Présentation au 90e congrès de l'Acfas
le 9 mai 2023

La découverte en sciences humaines

Qu'est-ce qu'une preuve solide en sciences humaines?

La valeur p : solution simple à un problème complexe.

Décision binaire : être ou ne pas être $p < .05$.

<u>P-VALUE</u>	<u>INTERPRETATION</u>
0.001	HIGHLY SIGNIFICANT
0.01	
0.02	
0.03	
0.04	SIGNIFICANT
0.049	
0.050	OH CRAP. REDO CALCULATIONS.
0.051	ON THE EDGE OF SIGNIFICANCE
0.06	
0.07	HIGHLY SUGGESTIVE, SIGNIFICANT AT THE $P < 0.10$ LEVEL
0.08	
0.09	
0.099	HEY, LOOK AT THIS INTERESTING SUBGROUP ANALYSIS
≥ 0.1	

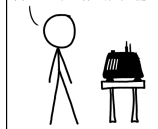
La découverte en sciences humaines

- La terre est ronde, $p < .05$.
- La pénicilline est un antibiotique, $p < .05$.
- $E = mc^2$, $p < .05$.
- Le conditionnement opérant, $p < .05$.
- L'obéissance à l'autorité, $p < .05$.
- *Le soleil a explosé*, $p < .05$.



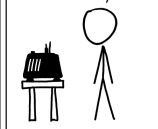
FREQUENTIST STATISTICIAN:

THE PROBABILITY OF THIS RESULT HAPPENING BY CHANCE IS $\frac{1}{36} = .027$. SINCE $p < 0.05$, I CONCLUDE THAT THE SUN HAS EXPLODED.



BAYESIAN STATISTICIAN:

BET YOU \$50 IT HASN'T.



Les problèmes

L'hypothèse nulle n'est jamais susceptible d'être vraie.

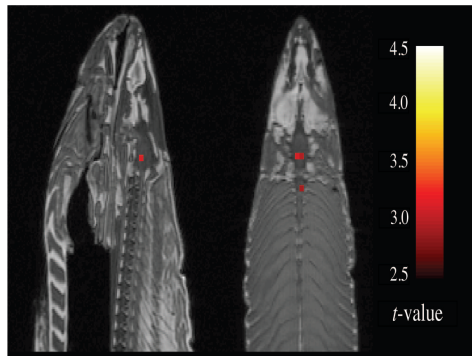
La significativité statistique n'implique pas la significativité pratique.

Une valeur- p ne garantit pas la reproductibilité.

L'obsession de la valeur- p aggrave la fraude scientifique.

Une saine utilisation de la valeur- p n'aide en rien.

Quelques cas : Le saumon de Bennett et al. (2010)



Saumon décédé sous IRMf montrant dans une tâche de cognitions sociales. Tiré de Bennett et al. (2010), p. 4.

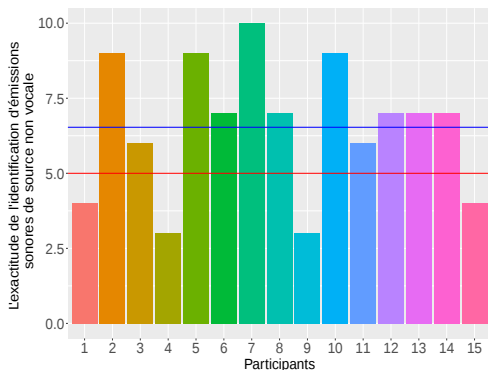
Quelques cas : Les émissions sonores de Pichette (2015)

Objectif: Examiner la perception de l'accent dans des émissions sonores de source non vocale.

Quinze adultes francophones sont soumis à des tests de différentiation perceptuelle de 10 émissions éructatives et flatulentiellles présentées aléatoirement.

L'hypothèse : les marqueurs dialectaux et non-natifs soient moindres pour ces émissions sonores.

Quelques cas : Les émissions sonores de Pichette (2015)



Le résultat : $\chi^2(1) = 14.107, p = 1.7 \times 10^{-4}$.

Pour 98 identifications exacts et 52 identifications incorrects.

Quelques cas : L'étude de McShane & Gal (2016)

Cinquante-deux pour cent (52%) des sujets ayant pris le médicament A ont guéri, tandis que quarante-quatre pour cent (44%) des sujets ayant pris le médicament B ont guéri.

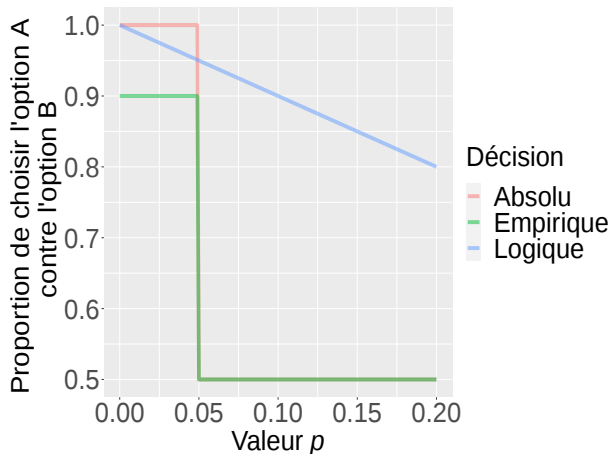
Un test de l'hypothèse nulle selon laquelle il n'y a pas de différence entre le médicament A et le médicament B en termes de probabilité de guérison donne une valeur p de x .

A. Je conseille le médicament A.

B. Je conseille le médicament B.

C. Je dirais qu'il n'y a pas de différence entre le médicament A et le médicament B.

Quelques cas : L'étude de McShane & Gal (2016)



Prouver en sciences humaines

La valeur p est innocente.

La vraie coupable est la difficulté de *prouver* en sciences humaines.

Prouver en sciences humaines

Exacerbation du fardeau de la preuve

- suffisamment de puissance pour détecter un résultat significatif?
- l'hypothèse nulle est-elle contenue dans l'intervalle de confiance?
- les indices d'ajustement sont-ils adéquats?

Prouver en sciences humaines

Ce qui maintient l'utilisation de la valeur p :

- solution simple, facile, générale et conventionnelle (Béland et al., soumis);
- l'expertise approfondie d'un domaine extérieur de celui de la spécialisation initiale, et;
- l'apprentissage plus difficile (Cousineau & Harding, 2017) et l'enseignement lui-même.

Prouver en sciences humaines

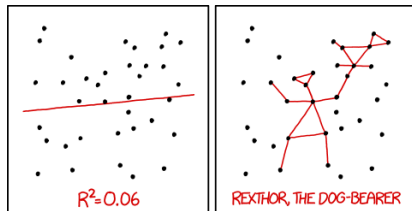
Un chercheur s'engage *rarement* à ne pas trouver (Howell, 2008).

La valeur p nous en apprend plus sur le chercheur que sur ses données.

Solutions

Miser sur ...

- une meilleure compréhension et l'utilisation des tailles d'effet;
- les représentations graphiques (les statistiques sont en appui), et;
- un meilleur enseignement des tests d'hypothèses.



I DON'T TRUST LINEAR REGRESSIONS WHEN IT'S HARDER TO GUESS THE DIRECTION OF THE CORRELATION FROM THE SCATTER PLOT THAN TO FIND NEW CONSTELLATIONS ON IT.

Conclusions

Un outil statistique n'est efficace qu'avec de bonnes données soutenues par une bonne expérimentation.

Ultimement, la reproduction d'étude (sans binariser non plus ce processus).

Références I

Béland, S., Cantinotti, M., Cousineau, D., Barroca-Paccard, M., Boislard-Pépin, M.-A., Bourrassa, C., Bureau, J., Caron, P.-O., Chanquoy, L., Chénier, C., Daoust, J.-F., Dionne, É., Forté, S., Frenette, É., Girard, S., Grenon, V., Harding, B., Lacourse, É., Lalande, D., ... Vugt, F. van. (soumis).

Pourquoi tout ce bruit autour de la valeur p ?

Bennett, C. M., Baird, A. A., Miller, M. B., & Wolford, G. L. (2010). Neural correlates of interspecies perspectives taking in the post-moterm atlantic salmon : An argument for proper multiple comparisons correction. *Journal of the Serendipitous and Unexpected Results*, 1(1), 1–5.

Références II

- Cousineau, D., & Harding, B. (2017). Pourquoi les statistiques sont-elles difficiles à enseigner et à comprendre? Quelques réflexions. *Revue de Psychoéducation*, 46(2), 397–419.
<https://doi.org/https://doi.org/10.7202/1042257ar>
- Howell, D. C. (2008). *Méthodes statistiques en sciences humaines*. de Boeck.
- McShane, B. B., & Gal, D. (2016). Blinding us to the obvious? The effect of statistical training on the evaluation of evidence. *Management Science*, 62(6), 1707–1718.
<https://doi.org/10.1287/mnsc.2015.2212>
- Pichette, F. (2015). *Perception des caractéristiques phonologiques non natives par des adultes francophones*. Communication présentée au 82e congrès de l'ACFAS, Montréal.