

Bayesian statistics

Null Hypothesis Significance Testing

Khương Quỳnh Long

Hà Nội, 08/2019

<https://gitlab.com/LongKhuong/adhere-bayesian-statistics>

- ▶ Nghiên cứu RCT so sánh hiệu quả của thuốc A trong điều trị viêm dạ dày
- ▶ 400 bệnh nhân được phân ngẫu nhiên vào 2 nhóm (mỗi nhóm 200 bệnh nhân) 1 nhóm sử dụng thuốc A và 1 nhóm sử dụng giả dược
- ▶ Kết quả:
 - Trong nhóm chứng: 70% bệnh nhân khỏi bệnh
 - Nhóm thuốc A: 80% bệnh nhân khỏi bệnh
 - $\text{Probtest}_{A \text{ vs placebo}}: z = 2.31, p \text{ value} = 0.0209$
- ▶ Giải thích kết quả?
- ▶ P value là gì?
- ▶ “Có ý nghĩa thống kê” là gì?

Nội dung

1

- Test of Significance

2

- Test of Hypothesis

3

- Null Hypothesis Significance Testing

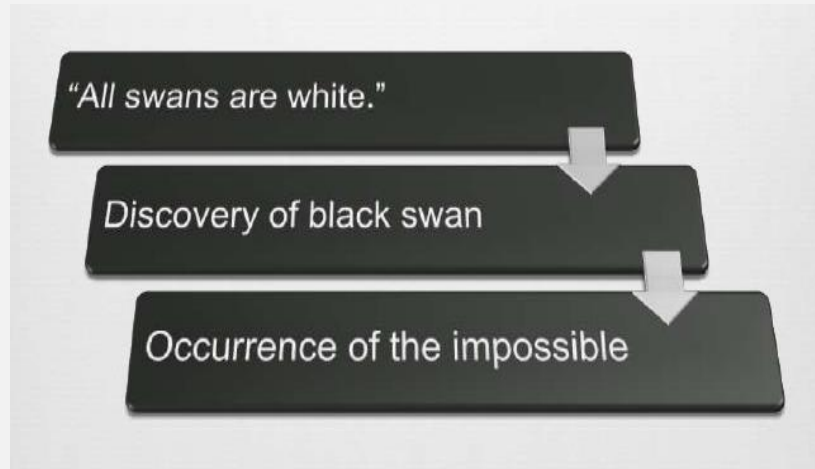
4

- P value và những hiểu lầm

1. Test of Significance

Test of Significance

- ▶ Được giới thiệu bởi Ronald Fisher vào 1920s
- ▶ Dựa trên triết lý phản nghiệm (falsificationism)
 - *Không chứng minh được một giả thuyết*
 - *Chỉ có thể bác bỏ giả thuyết*



Test of Significance

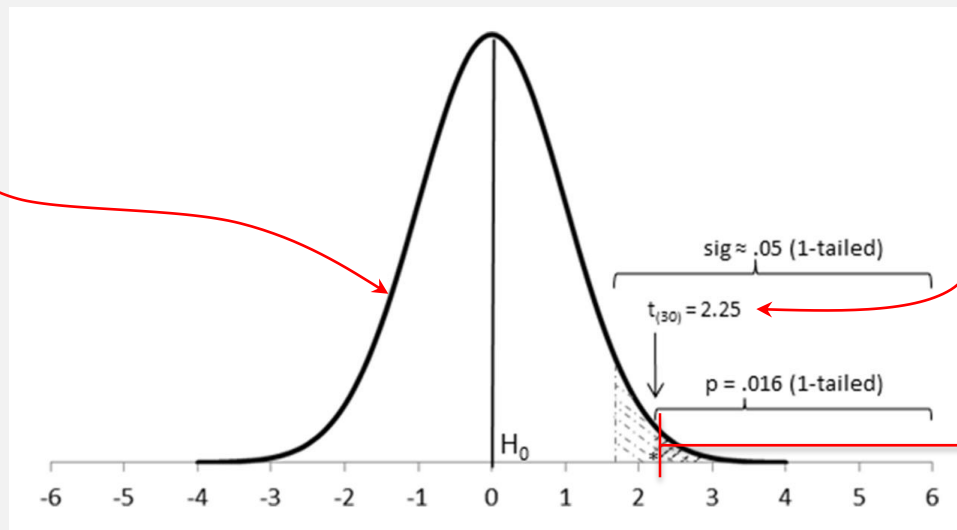
- ▶ B1. Lựa chọn phép kiểm phù hợp
- ▶ B2. Đặt một giả thuyết vô hiệu (Null) H_0
- ▶ B3. Tính xác suất quan sát được biến cố và những trường hợp “hiếm hơn” (observed and more extreme results) khi H_0 đúng $P(D|H_0)$ hay P value

Test of Significance

► P-value

Ví dụ: PP của sự khác biệt
~ Student t (0, 1, df = 30)
i.e., $t \sim (\text{khác biệt})/\text{se}$

Giả định: $H_0 = 0$



Giá trị thống kê
 $t_{(30)} = 2.25$

P-value: **nếu H_0 đúng** (không có khác biệt giữa 2 nhóm, tức phân phối trên xảy ra), xác suất để quan sát được kết quả như đang có ($t_{(30)} = 2.25$) và những kết quả hiếm hơn ($t > 2.25$) là 0.016 (p 1 đuôi)

P value là xác suất **cộng gộp** (cumulative probability), không phải xác suất tại 1 điểm (point probability)

Test of Significance

- ▶ B4. Đánh giá kết quả “có ý nghĩa thống kê”: P value **càng nhỏ** → giả thuyết H_0 **càng ít khả năng xảy ra** → càng cho thấy bằng chứng bác bỏ H_0
- ▶ Trình bày đầy đủ P value rất quan trọng (không nên trình bày <0.05 , $<0.01\dots$)

Note: Các bước có thể tiến hành khi dữ liệu đã có sẵn (không cần trong giai đoạn lập kế hoạch nghiên cứu)

Test of Significance

- ▶ Cách tiếp cận của Fisher linh hoạt
- ✓ Mức ý nghĩa (significance level) tùy thuộc vào bối cảnh nghiên cứu
- ✓ Không cứng nhắc (vd. P value = 0.049 và 0.051 có cùng mức ý nghĩa thống kê quanh mức 0.05)
- ▶ Cách tiếp cận **1 chiều** (chỉ có giả thuyết H_0) → bác bỏ H_0 không nhất thiết là ủng hộ giả thuyết đối lập

2. Test of Hypothesis

Test of Hypothesis

- ▶ Jerzy Neyman và Egon Pearson giới thiệu một cách tiếp cận khác (1928)
- ▶ Khác Fisher, nhà thống kê phải có kế hoạch phân tích trong giai đoạn đầu của nghiên cứu

Test of Hypothesis

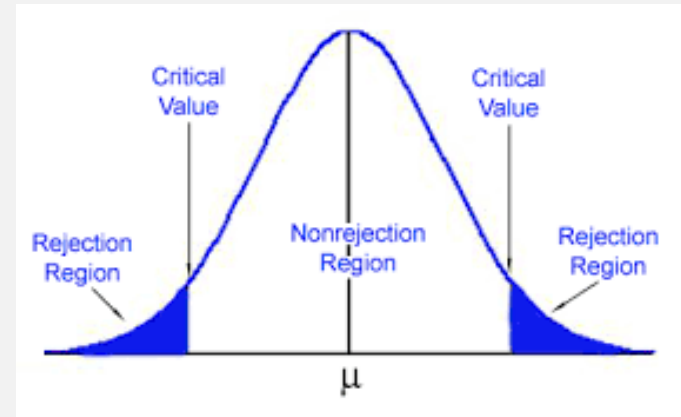
- ▶ B1. Lựa chọn phép kiểm phù hợp
- ▶ B2. Đặt giả thuyết chính (main hypothesis) $H_M \neq H_0$
- ▶ B3. Đặt giả thuyết đảo (alternative hypothesis) H_A
- ▶ B4. Cân nhắc xác suất sai lầm loại I và loại II
- ▶ B5. Tính cỡ mẫu để đạt power $(1 - \beta)$ tốt
- ▶ B6. Tính toán vùng bác bỏ (rejection region)
- ▶ B7. Tính toán giá trị thống kê (test value)
- ▶ B8. Quyết định ủng hộ giả thuyết chính hay đảo

Trước
khi thu
thập số
liệu

Test of Hypothesis

- ▶ Nếu giá trị thống kê (test value) **nằm trong** vùng bác bỏ (rejection region) → bác bỏ H_M → ủng hộ H_A
- ▶ Nếu **nằm ngoài** vùng bác bỏ và **đủ power** → chấp nhận H_M
- ▶ Nếu **nằm ngoài** vùng bác bỏ và **không đủ power** → Không kết luận được gì

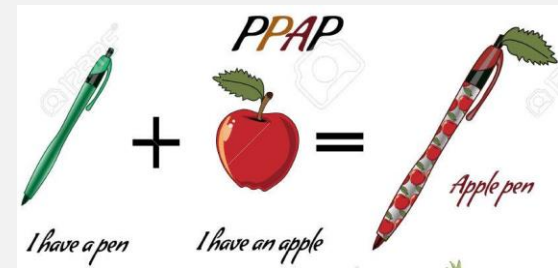
Note: **Không có P value**



3. Null Hypothesis Significance Testing

Null Hypothesis Significance Testing

- ▶ NHST là quy trình được sử dụng phổ biến nhất hiện nay
- ▶ Là sự kết hợp giữa quy trình kiểm định giả thuyết của Neyman-Pearson và triết lý phản nghiệm của Fisher (tuy nhiên chưa thực sự hòa hợp)
- ✓ Quy trình tương tự như test of hypothesis
- ✓ Khi $P \text{ value} < \alpha$, bác bỏ $H_0 \rightarrow$ chấp nhận H_A , nếu không thì chấp nhận H_0



4. Các vấn đề của P value

Table 1 Twelve P-Value Misconceptions

1	<i>If $P = .05$, the null hypothesis has only a 5% chance of being true.</i>
2	<i>A nonsignificant difference (eg, $P \geq .05$) means there is no difference between groups.</i>
3	<i>A statistically significant finding is clinically important.</i>
4	<i>Studies with P values on opposite sides of .05 are conflicting.</i>
5	<i>Studies with the same P value provide the same evidence against the null hypothesis.</i>
6	<i>$P = .05$ means that we have observed data that would occur only 5% of the time under the null hypothesis.</i>
7	<i>$P = .05$ and $P \leq .05$ mean the same thing.</i>
8	<i>P values are properly written as inequalities (eg, "$P \leq .02$" when $P = .015$)</i>
9	<i>$P = .05$ means that if you reject the null hypothesis, the probability of a type I error is only 5%.</i>
10	<i>With a $P = .05$ threshold for significance, the chance of a type I error will be 5%.</i>
11	<i>You should use a one-sided P value when you don't care about a result in one direction, or a difference in that direction is impossible.</i>
12	<i>A scientific conclusion or treatment policy should be based on whether or not the P value is significant.</i>

Goodman, S. (2008). A Dirty Dozen: Twelve P-Value Misconceptions. *Seminars in Hematology*, 45(3), 135–140.

1. Vấn đề logic

- ▶ P value là xác suất quan sát được dữ liệu (observed and more extreme outcomes) khi H_0 đúng
- ▶ Vì H_0 được giả định là đúng $\rightarrow H_0$ không thể sai, hay không thể bác bỏ H_0
- ▶ Nếu H_0 sai (bác bỏ H_0) \rightarrow toàn bộ phép kiểm bị vi phạm giả định
- ▶ Mặt khác, H_0 đúng nên không cần chứng minh ($P > \text{mức ý nghĩa}$)

1. Vấn đề logic

- ▶ P value cho biết mức độ khả dĩ của dữ liệu $P(D|H)$ chứ không cho biết mức độ khả dĩ của giả thuyết
- ▶ Trong thực tế, nhà nghiên cứu cần biết $P(H|D)$
- ➔ Chỉ số P không đáp ứng được nhu cầu này

2. Kiểm định nhiều giả thuyết

- ▶ Mỗi lần bác bỏ $H_0 \rightarrow H_1$, chấp nhận 1 sai sót (dương tính giả) là 5%, hay xác suất đúng là 95%

Ví dụ: so sánh đường huyết 3 thời điểm: t_1 , t_2 , t_3

- So sánh t_1 với t_2 , t_1 với t_3 , t_2 với t_3
- Xác suất **cả 3 giả thuyết đúng** = $0.95 \cdot 0.95 \cdot 0.95 = 0.86$
- Xác suất **(+) giả ít nhất 1 lần** = $1 - 0.86 = 0.14 = 14\%!!$

- ▶ Hiệu chỉnh Bonferroni

$P < 0.05/(\text{tổng số giả thuyết})$ thì mới “có ý nghĩa thống kê”

3. Phụ thuộc vào cỡ mẫu

- ▶ Cỡ mẫu càng lớn càng dễ đạt được P value có “ý nghĩa thống kê”

Ví dụ:

- Tung đồng xu 10 lần: 6 lần sấp, 4 lần ngửa ($X_s \text{ sấp} = 0.6$)
 - ✓ H_0 : đồng xu cân bằng ($x_s \text{ sấp} = \text{ngửa} = 0.5$)
 - ✓ P value = $P(\geq 6|0.5) = 0.377 \rightarrow$ đồng xu cân bằng
- Tung đồng xu 100 lần: 60 lần sấp, 40 lần ngửa ($X_s \text{ sấp} = 0.6$)
 - ✓ H_0 : đồng xu cân bằng
 - ✓ P value = $P(\geq 60|0.5) = 0.028 \rightarrow$ đồng xu không cân bằng

3. Phụ thuộc vào cỡ mẫu

Ý nghĩa thống kê vs. Ý nghĩa thực tế
(Statistical significance vs. Practical significance)

Sự khác biệt giữa “có ý nghĩa thống kê” & “không có ý nghĩa thống kê” là không có ý nghĩa thống kê!¹

1. McShane BB, Gal D, Gelman A, Robert C, Tackett JL. Abandon Statistical Significance. The American Statistician. 2019;73(sup1):235-45.

4. Ý định dừng thử nghiệm

(intention stopping rule)

- ▶ Mục tiêu nghiên cứu: điều tra hiệu quả thuốc A và B

- ▶ Thử nghiệm 1:

Ý định: thử nghiệm trên 6 mẫu bệnh nhân: đếm được 5 mẫu thuốc A > B, 1 mẫu B > A

- ▶ Thử nghiệm 2:

Ý định: thử nghiệm đến khi nào quan sát thấy 1 lần thuốc B > A thì ngừng thử nghiệm:

Lần thứ 6 quan sát được thuốc B > A → ngừng (cũng 5 mẫu thuốc A > B, 1 mẫu B > A)

4. Ý định dừng thử nghiệm

(intention stopping rule)

- ▶ P value = P(quan sát kết cuộc + kết cuộc hiếm hơn)
- ▶ Giả thuyết H_0 : hiệu quả 2 thuốc tương đồng: $P(A > B) = \frac{1}{2}$
- ▶ Giả thuyết H_1 : thuốc A > B (sử dụng P 1 đuôi)
- ▶ Thử nghiệm 1:

$$P(\text{quan sát kết cuộc}) = 6 * (\frac{1}{2}) * (\frac{1}{2})^5$$


$$P(\text{kết cuộc hiếm hơn}) = (\frac{1}{2})^6$$

$P = 0.11 \rightarrow$ Chấp nhận H_0
Thuốc A \approx B 

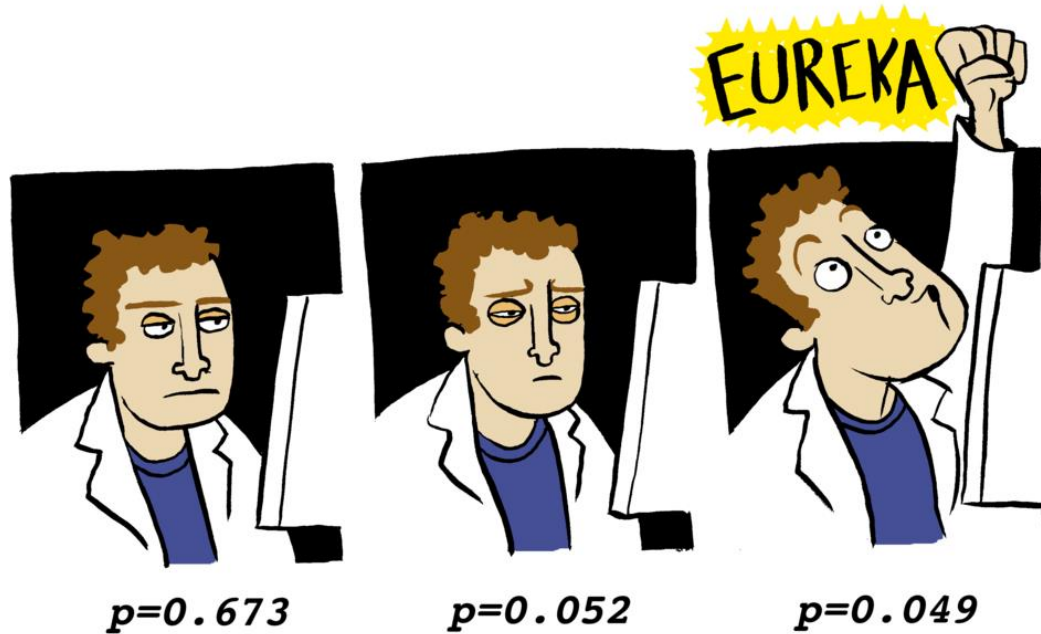
- ▶ Thử nghiệm 2:

$$P(\text{quan sát kết cuộc}) = (\frac{1}{2})^5 * (\frac{1}{2})$$

$$P(\text{kết cuộc hiếm hơn}) = (\frac{1}{2})^6$$

$P = 0.03 \rightarrow$ Bác bỏ H_0
Thuốc A > B 

5. “P hacking”, “Fishing”



5. “P hacking”, “Fishing”,

- Retire statistical significance

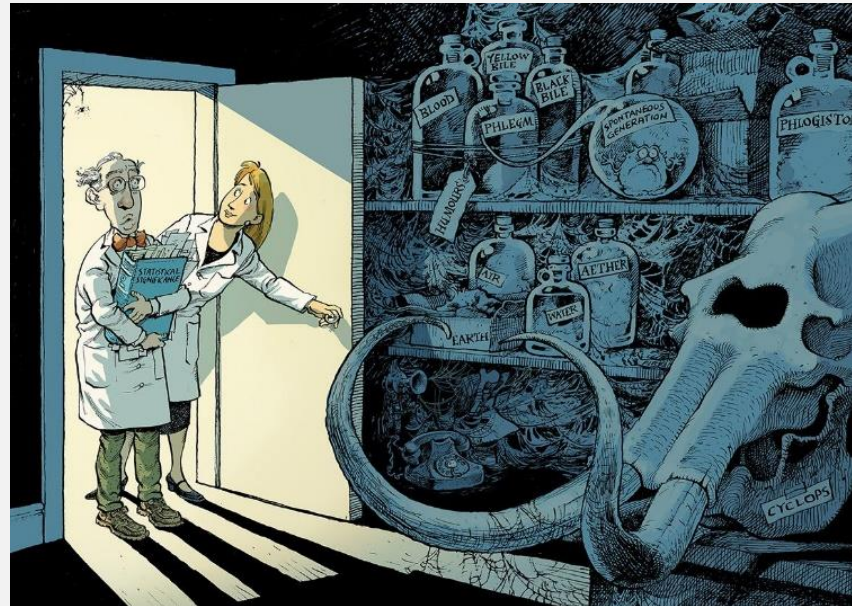
<https://www.nature.com/articles/d41586-019-00857-9>

- The American Statistician – Statistical inference in the 21st century: a world beyond $p < 0.05$

<https://www.tandfonline.com/toc/utas20/73/sup1>

- Abandon Statistical Significance

<https://www.tandfonline.com/doi/full/10.1080/00031305.2018.1527253>



6. Không cân nhắc tiên định



- ▶ Tung cây đinh 10 lần, xuất hiện mặt “head” là 3 lần
- ▶ H_0 : cây đinh cân bằng giữa head và tail
- ▶ $P \text{ value} = P(H \leq 3 | 0.5) = 0.172$
- ➔ Chấp nhận $H_0 \rightarrow$ cây đinh cân bằng!!!

Tóm lại

- ▶ P value có nhiều điểm hạn chế
- ▶ Tuy nhiên, P value vẫn có giá trị của nó
- ▶ Cách sử dụng
 - Cách dạy thống kê
 - Cách áp dụng thống kê
- ▶ Ít nhất là hiểu được hạn chế khi sử dụng

Thank you !