

# Trị số p

Huỳnh Tấn Tài

Trong một cuộc họp hàng tháng, nhóm điều tra lương của nhân viên thư ký tại trường đại học mà bạn đang công tác đưa ra những số liệu sau đây:

	Nhân viên nam	Nhân viên nữ
Số lượng (sample size n)	30	36
Lương trung bình (sample mean) USD	27,233.33	25,855.81
Độ lệch chuẩn (standard deviation)	3,475.54	2,580.02

Căn cứ vào bảng số liệu trên, một số thành viên yêu cầu ban điều hành (administration) phải có những chính sách làm giảm sự sai biệt về lương giữa nhân viên nam và nhân viên nữ. Trong khi thảo luận, nhóm các nhà nghiên cứu y sinh học có nhắc tới trị số p (p-value) mà phần đông các thành viên khác tại buổi họp hiểu rất là lơ mơ. Là một thành viên trong ban điều hành, bạn tin rằng không có sự khác biệt về lương bổng giữa nhân viên nam và nhân viên nữ. Vậy câu hỏi là ban điều hành phải trả lời ra sao cho các thành viên trong cuộc họp.

Trong thống kê học cổ điển (còn được gọi là thống kê học tần suất – frequentialism), để tiến hành thử nghiệm giả thuyết (hypothesis testing), có hai phương cách cơ bản:

1. Dùng một độ ý nghĩa (level of significance) định trước, xác định trị số ngưỡng (critical values), rồi xem kê số thử nghiệm (test statistic) có nằm trong vùng bác bỏ (rejection region) hay không. Điều này giống như trong thi nhảy cao, xà ngang được kê trước và thí sinh phải rúng nhảy qua xà ngang này. Đây là phương pháp của trường phái Neyman-Pearson.
2. Xác định độ ý nghĩa chính xác tương ứng với kê số thử nghiệm. Điều này cũng giống như chuyện thí sinh cố gắng nhảy cao hết sức mình mà không có xà ngang, rồi sau đó ban giám khảo mới bảo thí sinh là đã nhảy cao bao nhiêu, ở mức xà ngang nào. Đây là phương pháp của trường phái Fisher.

Cách tiến hành theo phương cách thứ nhất như sau (Hình 1):

## **Giả thuyết không (null hypothesis) và giả thuyết khác (alternate hypothesis)**

$H_0: \mu (x\text{-bar } 1 - x\text{-bar } 2) = 0$  Lương giữa nhân viên nam và nữ như nhau

$H_1: \mu (x\text{-bar } 1 - x\text{-bar } 2) \neq 0$  Lương giữa nhân viên nam và nữ khác nhau

## **Kê số thử nghiệm (test statistic):**

$t = 1.7971$

## **Độ ý nghĩa (level of significance):**

$\alpha = 0.05$

**Trị số ngưỡng (critical values):**

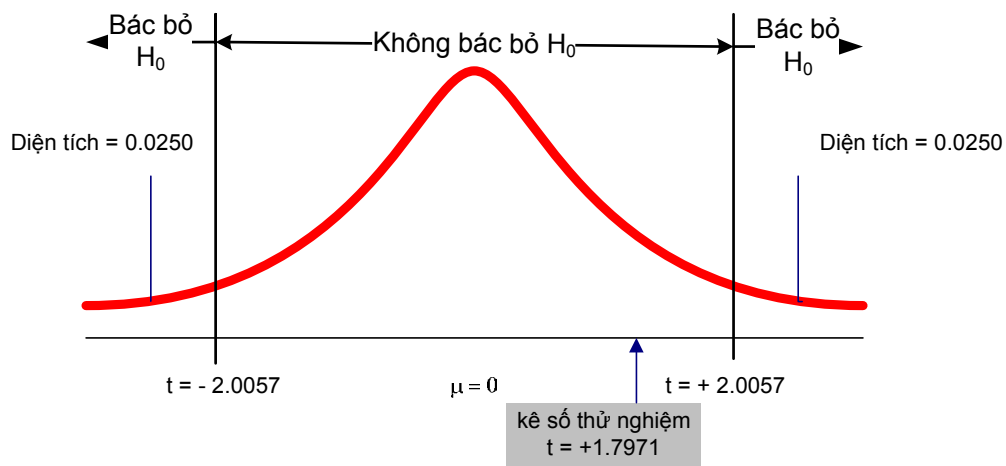
$$t = \pm 2.0057$$

**Nghiệm pháp quyết định (Decision rule):**

Bác bỏ  $H_0$  nếu thử nghiệm  $< -2.0057$  hoặc  $> +2.0057$ ; nếu không, đừng bác bỏ  $H_0$ .

**Kết luận:**

Không thể bác bỏ ý kiến lưỡng của nhân viên nam và nữ không có khác nhau (Đây là ý kiến của ban điều hành).

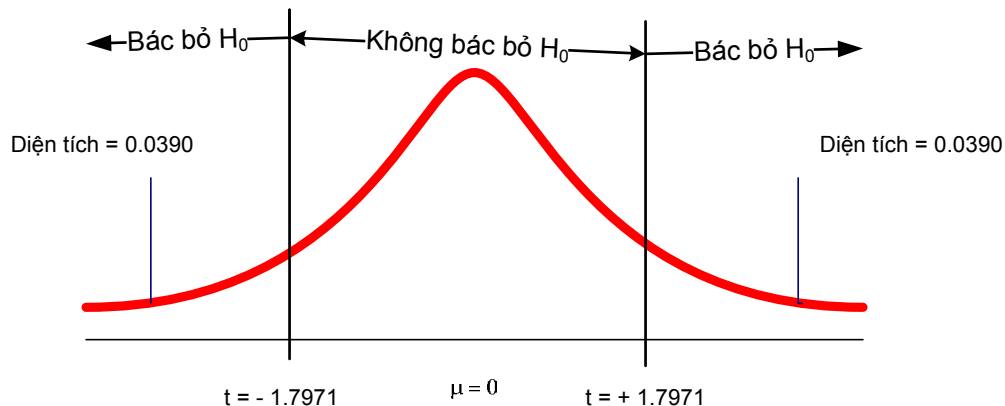


Hình 1

Kê số thực nghiệm  $t$  được tính để xem các số liệu của mẫu cách *giả thuyết không*  $H_0$  bao xa. Nếu  $t$  thực nghiệm càng xa (extreme), theo hướng của *giả thuyết khác*  $H_1$  thì có nghĩa là ít có khả năng, ít có cơ hội để cho *giả thuyết không* được xem là đúng. Các khái niệm “ít có khả năng, ít có cơ hội” được diễn giải bằng khái niệm trị số  $p$ . Trị số  $p$  trong thí dụ trên đây bằng 0.0780.

Trị số  $p$  cũng có thể xem như là độ ý nghĩa  $\alpha$  thấp nhất mà *giả thuyết không* có thể bị bác bỏ (Hình 2). Ở mức độ đó, trị số của  $t$  thực nghiệm bằng trị số ngưỡng. Với  $p = 0.0780$ , nếu  $\alpha$  lớn hơn  $p$ , chẳng hạn 0.1, thì *giả thuyết không* sẽ không thể bị bác bỏ. Ngược lại nếu  $\alpha$  nhỏ hơn trị số  $p$ , thí dụ 0.05 hoặc 0.01, thì *giả thuyết không* bị bác bỏ. Trở lại thí dụ nhảy cao, với tất cả kinh nghiệm và thời gian thực tập (số liệu từ mẫu), thí sinh cố gắng hết mức nhảy cho thật cao (test statistic value). Ở mức độ cao đó, thí sinh được cho một đẳng cấp ( $p$ -value). Còn chuyện thí sinh có thắng giải hay không còn tùy thuộc vào xà ngang (level of significance  $\alpha$ ).

Ngày nay với các chương trình chuyên về thống kê như Minitab, SPSS, SAS hoặc các chương trình doanh nghiệp MS Excel, trị số p có thể được tính một cách chính xác. Trị số p, khi được tính chính xác bằng các chương trình trên, cho nhiều thông tin hơn là một câu phát biểu về ý nghĩa thống kê vì bạn có thể đánh giá ý nghĩa ở bất cứ độ ý nghĩa nào mà bạn chọn.



Hình 2

Sau khi đã tiến hành cả hai phương cách trên, bạn rút ra kết luận sau đây:

1. Nếu trị số p nhỏ hơn độ ý nghĩa  $\alpha$ , bạn có thể bác bỏ *giả thuyết không*  $H_0$ . Nói cách khác, số liệu có ý nghĩa thống kê ở độ ý nghĩa  $\alpha$  (data are statistically significant at level  $\alpha$ ).
2. Nếu trị số p bằng hoặc lớn hơn độ ý nghĩa  $\alpha$ , bạn không thể bác bỏ *giả thuyết không*  $H_0$ . Bạn biết rằng không thể bác bỏ không có nghĩa là bạn chấp nhận giả thuyết  $H_0$ .

Bạn gửi điện thư cho tất cả thành viên của cuộc họp, giải thích lý do tại sao bạn điều hành vẫn tin rằng không có sự khác biệt lương giữa nam nhân viên và nữ nhân viên. Tất cả đều đồng ý tạm thời xếp lại vấn đề này. Riêng bản thân bạn học được một điều bổ ích: trị số p không chỉ dành cho nghiên cứu khoa học mà còn hữu ích trong công việc thường ngày của bạn. Một trong những đồng nghiệp y sinh học còn hứa giảng dạy thêm cho bạn về thống kê học Bayes và thống kê học Markov trong thời gian tới.

### Tài liệu tham khảo

1. Weirs, RD: Introduction to Business Statistics. 3<sup>rd</sup> ed. Duxbury Press. Pacific Grove, CA. 1998.
2. Moore, DS: The Basic Practice of Statistics. 3<sup>rd</sup> ed. Freeman. New York, NY. 2004.
3. Neufeld, JL: Learning Business Statistics with Microsoft Excel 2000. Prentice Hall. Upper Saddle River, NJ. 2001.

4. Nguyễn Văn Tuấn: Ý nghĩa của trị số P (P-Value) trong nghiên cứu khoa học. <http://www.ykhoanet.com/> tiếp cận ngày 28 tháng 11 năm 2005.
5. Goodman, S: Toward Evidence-Based Medical Statistics: 1. The P Value Fallacy. *Ann Intern Med.* 1999; 130:995-1004.
6. Goodman, S: Toward Evidence-Based Medical Statistics: 2. The Bayes Factor. *Ann Intern Med.* 1999; 130:1005-1013.