California Polytechnic State University, San Luis Obispo

JAY L. DEVORE

Tài liệu môn học

Probability and Statistics for Engineering and Sciences

XÁC SUẤT THỐNG KÊ ỨNG DỤNG

Người dịch:
Nguyễn Hồng Nhung
Hoàng Thị Minh Thảo
Lê Thị Mai Trang
Nguyễn Ngọc Tứ

Bộ môn Toán - ĐH SPKT, Tp. Hồ Chí Minh - Năm 2017

Mục lục

MỞ ĐẦU

Chương 1

TỔNG QUAN VÀ THỐNG KÊ MÔ TẢ

Giới thiệu

Chương 2

PHÉP TÍNH XÁC SUẤT

Giới thiệu

2.1 Không gian mẫu và biến cố

2.2 Các tiên đề và tính chất của xác suất

2.3 Giải tích tổ hợp

2.4 Xác suất có điều kiện

2.5 Sự độc lập

Chương 3

BIẾN NGẪU NHIÊN RỜI RẠC VÀ PHÂN PHỐI XÁC SUẤT

Giới thiệu

3.1 Biến ngẫu nhiên

3.2 Phân phối xác suất của biến ngẫu nhiên rời rạc

3.3 Kỳ vọng và phương sai

3.4 Phân phối nhị thức

3.5 Phân phối nhị thức âm và siêu bội

3.6 Phân phối Poisson

Chương 4

BIẾN NGẪU NHIÊN LIÊN TỤC VÀ PHÂN PHỐI XÁC SUẤT

Chương 5

PHÂN PHỐI XÁC SUẤT ĐỒNG THỜI VÀ MẪU NGẪU NHIÊN

Chương 6

ƯỚC LƯỢNG ĐIỂM

Giới thiệu

6.1 Một số khái niệm tổng quát về ước lượng điểm

6.2 Các phương pháp ước lượng điểm

Chương 7

ƯỚC LƯỢNG KHOẢNG

Giới thiệu

7.1 Các tính chất cơ bản của khoảng tin cậy

7.2 Khoảng tin cậy mẫu lớn cho trung bình tổng thể

7.3 Các khoảng dựa trên phân phối chuẩn

7.4 Khoảng tin cậy của phương sai và độ lệch chuẩn của phân phối chuẩn

Chương 8

KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ

Giới thiệu

8.1 Giả thiết và thủ tục kiểm định

8.2 Kiểm định về trung bình tổng thể

8.3 Kiểm định về tỷ lệ

8.4 P giá trị

8.5 Một số chú ý về chọn thủ tục kiểm định

Chương 9

CÁC KẾT LUẬN DỰA TRÊN HAI MẪU

Giới thiệu

Chương 7 và 8 đã trình bày về khoảng tin cậy và các bước kiểm định giả thuyết cho giá trị trung bình μ, tỷ lệ p, và phương sai σ^2. Ở đây chúng ta mở rộng các phương pháp cho trung bình, tỷ lệ, và phương sai của hai tổng thể khác nhau. Ví dụ, cho μ_1 kí hiệu độ cứng trung bình Rockwell thực sự đối với mẫu thép được xử lý nhiệt và μ_2 là độ cứng trung bình thực sự đối với mẫu thép đã qua cán nguội. Thì người điều tra có thể muốn sử dụng các mẫu quan sát độ cứng từ mỗi loại thép làm cơ sở để ước lượng khoảng $\mu_1 - \mu_2$ (hiệu giữa hai độ cứng). Ví dụ khác, cho p_1 kí hiệu cho tỷ lệ thực sự của các pin nickel-cadmium được sản xuất trong điều kiện hiện tại mà có tuổi thọ ngắn và p_2 kí hiệu cho tỷ lệ thực của pin bị lỗi trên nhưng được sản xuất trong điều kiện sản xuất đã thay đổi. Nếu nhờ vào điều kiện sản xuất sửa đổi mà làm giảm được tỷ lệ các pin lỗi, một kỹ sư quản lý chất lượng sẽ muốn sử dụng thông tin mẫu để kiểm định giả thuyết ban đầu $H_0 : p_1 - p_2 = 0$ (tức là $p_1 = p_2$) so với giả thuyết đối, $H_a : p_1 - p_2 > 0 (p_1 > p_2)$.

9.1 Kiểm định z và khoảng tin cậy cho hiệu giữa hai trung bình

Các kết luận được đề cập trong phần này liên quan đến $\mu_1 - \mu_2$ là hiệu của trung bình hai mẫu khác nhau. Một nhà điều tra có thể muốn thử nghiệm giả thuyết về hiệu giữa khả năng chịu lực trung bình của hai loại bìa các tông. Một giả thuyết như vậy sẽ chỉ ra rằng $\mu_1 - \mu_2 = 0$ nghĩa là $\mu_1 = \mu_2$. Ngoài ra, có thể ước lượng $\mu_1 - \mu_2$ bằng cách tính khoảng tin cậy 95 %. Những kết luận như vậy đòi hỏi phải có được một mẫu các quan sát độ bền cho từng loại tấm bìa các tông.

Giả Thiết 9.1.1.

1. *$X_1, X_2, ..., X_m$ là một mẫu ngẫu nhiên từ một phân phối với giá trị trung bình là μ_1 và phương sai σ_1^2.*
2. *$Y_1, Y_2, ..., Y_m$ là một mẫu ngẫu nhiên từ một phân phối với giá trị trung bình là μ_2 và phương sai σ_2^2.*
3. *Mẫu X và Y độc lập với nhau.*

Việc sử dụng m cho số lần quan sát trong mẫu thứ nhất và n cho số lần quan sát trong mẫu thứ hai, cho phép hai cỡ mẫu khác nhau. Đôi khi điều này là vì việc lấy mẫu của một tổng thể này khó khăn hơn hoặc đắt tiền so với một tổng thể khác. Một số tình huống khác mà kích cỡ mẫu được lấy là bằng nhau ngay từ đầu, nhưng vì lý do vượt quá phạm vi thử nghiệm, mà kích thước mẫu thực tế có thể khác nhau. Ví dụ, bài tóm tắt của bài báo "A Randomized Controlled Trial Assessing the Effectiveness of Professional Oral Care by Dental Hygienists " (Intl. Of Dental Hygiene, 2008: 63-67) cho biết: "Bốn mươi bệnh nhân được phân ngẫu nhiên vào nhóm POC ($m = 20$) hoặc nhóm đối chứng ($n = 20$) . Một bệnh nhân trong nhóm POC và ba trong nhóm đối chứng đã bị loại do bệnh nặng hơn hoặc tử vong ". Dữ liệu được phân tích dựa trên và $m = 19$ và $n = 16$.
$\overline{X} - \overline{Y}$ là ước lượng tự nhiên của $\mu_1 - \mu_2$, là sự khác biệt giữa trung bình các mẫu tương ứng. Các quy trình kết luận được dựa trên chuẩn hóa của ước lượng tự nhiên này, vì vậy chúng ta cần biểu thức cho giá trị kỳ vọng và độ lệch chuẩn của $\overline{X} - \overline{Y}$.

Mệnh đề 9.1.2. *Giá trị kỳ vọng của $\overline{X} - \overline{Y}$ là $\mu_1 - \mu_2$, cho nên $\overline{X} - \overline{Y}$ là một ước lượng không chệch của $\mu_1 - \mu_2$. Độ lệch chuẩn của $\overline{X} - \overline{Y}$ là:*

$$\sigma_{\overline{X}-\overline{Y}} = \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$$

Nếu chúng ta coi $\mu_1 - \mu_2$ là một tham số θ, thì ước lượng của nó là $\hat{\theta} = \overline{X} - \overline{Y}$ với độ lệch chuẩn $\sigma_{\hat{\theta}}$ được tính như mệnh đề trên. Với giá trị của σ_1^2 và σ_2^2 đều đã biết, thì giá trị của độ lệch chuẩn này có thể được tính như trên. Khi σ_1^2 và σ_2^2 đều không biết trước, phương sai mẫu được dùng để ước lượng $\sigma_{\hat{\theta}}$.

9.1.1 Các bước kiểm định tổng thể chuẩn với phương sai đã biết

Trong chương 7 và 8, khoảng tin cậy và các bước kiểm định đầu tiên cho trung bình tổng thể μ dựa trên giả định rằng tổng thể là phân phối chuẩn với giá trị của phương sai tổng thể σ^2 đã được biết trước . Tương tự như vậy, trước tiên chúng ta giả sử rằng ở đây cả hai phân phối tổng thể là chuẩn và giá trị phương sai của cả hai là σ_1^2 và σ_2^2 được biết. Các tình huống trong đó một hoặc cả hai giả định này có thể được loại bỏ sẽ được trình bày ngắn gọn sau.
Bởi vì phân phối tổng thể là chuẩn, cả $\overline{X}$ và $\overline{Y}$ đều có phân phối chuẩn. Hơn nữa, sự độc lập của hai mẫu ngụ ý rằng hai trung bình mẫu cũng độc lập với nhau. Do đó, hiệu $\overline{X} - \overline{Y}$ cũng là phân phối chuẩn, với giá trị kỳ vọng $\mu_1 - \mu_2$ và độ lệch chuẩn $\sigma_{\overline{X}-\overline{Y}}$ được suy ra từ mệnh đề nêu trên. Chuẩn hóa $\overline{X} - \overline{Y}$ cho biến chuẩn hóa.

$$Z = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \tag{9.1}$$

Trong một bài toán kiểm định giả thuyết, giả thuyết ban đầu sẽ chỉ ra rằng $\mu_1 - \mu_2$ có một giá trị xác định. Gán giá trị giả thuyết này bằng Δ_0, chúng ta có $H_0 : \mu_1 - \mu_2 = \Delta_0$. Thông thường $\Delta_0 = 0$, trong trường hợp đó H_0 sẽ là $\mu_1 = \mu_2$. Một kết quả thống kê thử nghiệm từ việc thay thế $\mu_1 - \mu_2$ trong công thức (9.1) bằng giá trị Δ_0. Kiểm định thống kê Z thu được bằng cách chuẩn hóa $\overline{X} - \overline{Y}$ với giả định rằng H_0 là đúng, vì vậy trường hợp này nó có phân phối chuẩn chuẩn tắc. Kiểm định thống kê này có thể được viết là $(\hat{\theta} - \Delta_0)/\sigma_{\hat{\theta}}$, cùng dạng như một số kiểm định thống kê trong Chương 8.

Xem xét giả thuyết đối $H_a : \mu_1 - \mu_2 > \Delta_0$. Một giá trị $\overline{x} - \overline{y}$ được xem là lớn hơn Δ_0 (giá trị kỳ vọng của $\overline{X} - \overline{Y}$ khi H_0 là đúng) cung cấp dữ kiện chống lại H_0 và củng cố cho giả thuyết nghịch H_a. Giá trị $\overline{x} - \overline{y}$ tương ứng với z có giá trị dương và lớn. Do đó H_0 bị bác bỏ và giả thuyết nghịch H_a được chấp nhận nếu z lớn hơn hoặc bằng một giá trị tới hạn (critical value) được chọn thích hợp. Bởi vì kiểm định thống kê Z có phân phối chuẩn chuẩn tắc khi H_0 là đúng, miền bác bỏ phía bên phải $z \geq z_\alpha$ cho kiểm định với mức ý nghĩa α (xác suất của sai lầm loại I). Tương ứng vậy, miền bác bỏ của $H_a : \mu_1 - \mu_2 < \Delta_0$ và $H_a : \mu_1 - \mu_2 \neq \Delta_0$ thu được từ kiểm định với mức ý nghĩa α là phía bên trái và 2 phía.

Giả thuyết không $H_0 : \mu_1 - \mu_2 = \Delta_0$

Giá trị kiểm định thống kê: $z = \frac{\overline{x} - \overline{y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}$

Các giả thuyết đối:	Vùng bác bỏ với kiểm định α
$H_a : \mu_1 - \mu_2 > \Delta_0$	$z \geq z_\alpha$ (phía phải)
$H_a : \mu_1 - \mu_2 < \Delta_0$	$z \leq -z_\alpha$ (phía trái)
$H_a : \mu_1 - \mu_2 \neq \Delta_0$	$z \geq z_{\alpha/2}$ hoặc $z \leq -z_{\alpha/2}$ (2 phía)

Bởi vì đây là những kiểm định z, giá trị P-value đã được tính như cho kiểm định z ở chương 8 (ví dụ: giá trị P-value $= 1 - \Phi(z)$ cho bài kiểm định một phía phải).

Ví dụ 9.1 Phân tích một mẫu ngẫu nhiên gồm $m = 20$ mẫu thép cán lạnh để xem sức bền thì được sức bền trung bình $\overline{x} = 29.8$ ksi. Một mẫu ngẫu nhiên thứ hai gồm $n = 25$ mẫu thép được mạ kẽm 2 mặt có sức bền trung bình là $\overline{y} = 34.7ksi$. Giả sử rằng hai phân phối cho sức bền của 2 mẫu trên có phân phối chuẩn với $\sigma_1 = 4.0$ và $\sigma_2 = 5.0$ (số liệu được lấy từ bài báo "Zinc - Coated Sheet Steel" của Automotive Engr, 1984). Dữ liệu trên có chỉ ra được giá trị trung bình thật sự của hai sức bền trên μ_1 và μ_2 là khác nhau không? Hãy thực hiện kiểm định với mức ý nghĩa $\alpha = 0.01$

9.1.2 Dùng so sánh để xác định nguyên nhân

Các nhà điều tra thường quan tâm đến việc so sánh ảnh hưởng của hai phương pháp điều trị khác nhau dựa trên sự hồi đáp sau điều trị với hồi đáp khi không điều trị (điều trị và kiểm soát). Nếu các cá nhân hoặc đối tượng được sử dụng trong việc so sánh được chọn ngẫu nhiên vào hai nhóm điều kiện khác nhau, thì nghiên cứu này được xem là có tính quan sát. Khó khăn trong việc rút ra kết luận dựa trên một nghiên cứu có tính quan sát là mặc dù phân tích thống kê có thể chỉ ra sự khác biệt có ý nghĩa trong phản ứng giữa hai nhóm, nhưng sự khác biệt có thể là do một số yếu tố cơ bản chưa được kiểm soát chứ không phải sự khác nhau trong điều trị.

Ví dụ 9.2 Một lá thư trong Tạp chí của Hội y học Mỹ (1978) báo cáo rằng trong 215 bác sĩ nam tốt nghiệp từ Harvard thì chết giữa tháng 11 năm 1974 đến tháng 10 năm 1977, trong đó có 125 thực tập viên toàn thời gian sống trung bình 48.9 năm sau khi tốt nghiệp; trong khi đó có 90 người học liên kết sống trung bình 43.2 năm sau khi tốt nghiệp. Liệu dữ liệu cho rằng thời gian sống trung bình của thực tập viên toàn thời gian thì nhiều hơn thời gian sống trung bình của người học liên kết được không?

Một kết quả thử nghiệm được xem là ngẫu nhiên có kiểm soát khi các nhà nghiên cứu chỉ định các đối tượng cho hai phương pháp điều trị một cách ngẫu nhiên. Khi ý nghĩa thống kê được quan sát thấy trong một thí nghiệm như vậy, điều tra viên và các bên quan tâm khác sẽ tự tin hơn khi kết luận rằng sự khác biệt trong dấu hiệu hồi đáp là từ sự khác biệt trong cách điều trị. Một ví dụ rất nổi tiếng về loại thí nghiệm và kết luận này là thử nghiệm vaccine bệnh bại liệt Salk được mô tả trong Phần 9.4. Những vấn đề này được Moore và Freedman cùng các cộng sự thảo luận nhiều hơn trong các cuốn sách của họ (không phải sách về toán học) đã được liệt kê trong danh mục các tài liệu tham khảo ở Chương 1.

9.1.3 β và việc chọn kích cỡ mẫu

Xác suất của sai lầm loại II dễ dàng được tính toán khi cả hai mẫu đều là phân phối chuẩn với các giá trị σ_1 và σ_2 đã biết. Xem xét trường hợp giả thuyết đối $H_a : \mu_1 - \mu_2 > \Delta_0$. Đặt Δ' biểu thị cho giá trị của $\mu_1 - \mu_2$ mà lớn hơn hẳn Δ_0 (một giá trị mà H_0 là sai). Miền bác bỏ bên phải $z \geq z_\alpha$ có thể được viết lại dưới dạng $\overline{x} - \overline{y} \geq \Delta_0 + z_\alpha \sigma_{\overline{X}-\overline{Y}}$ .Như vậy.

$$\begin{aligned}\beta(\Delta') &= P(\text{không bác bỏ } H_0 \text{ khi mà } \mu_1 - \mu_2 = \Delta') \\ &= P(\overline{X} - \overline{Y} < \Delta_0 + z_\alpha \sigma_{\overline{X}-\overline{Y}} \text{ khi mà } \mu_1 - \mu_2 = \Delta')\end{aligned}$$

Khi $\mu_1 - \mu_2 = \Delta'$, $\overline{X} - \overline{Y}$ có phân phối chuẩn với giá trị trung bình Δ' và độ lệch chuẩn $\sigma_{\overline{X}-\overline{Y}}$ (cũng là độ lệch chuẩn khi H_0 đúng); sử dụng các giá trị này để chuẩn hóa bất đẳng thức trong ngoặc đơn cho ta xác suất mong muốn.

Giả thuyết đối	$\beta(\Delta') = P$(sai lầm loại II khi mà $\mu_1 - \mu_2 = \Delta'$)
$H_a : \mu_1 - \mu_2 > \Delta_0$	$\Phi(z_\alpha - \frac{\Delta'-\Delta_0}{\sigma})$
$H_a : \mu_1 - \mu_2 < \Delta_0$	$1 - \Phi(-z_\alpha - \frac{\Delta'-\Delta_0}{\sigma})$
$H_a : \mu_1 - \mu_2 \neq \Delta_0$	$\Phi(z_{\alpha/2} - \frac{\Delta'-\Delta_0}{\sigma}) - \Phi(-z_{\alpha/2} - \frac{\Delta'-\Delta_0}{\sigma})$
Trong đó $\sigma = \sigma_{\overline{X}-\overline{Y}} = \sqrt{(\sigma_1^2/m) + (\sigma_2^2/n)}$	

Như trong chương 8, mẫu kích cỡ m và n có thể được xác định để thỏa cả P(sai lầm loại I) = giá trị cụ thể α và P (sai lầm loại II khi $\mu_1 - \mu_2 = \Delta'$)= giá trị cụ thể β. Cho một kiểm định một phía phải, phương trình của biểu thức trước cho $\beta(\Delta')$ một giá trị cụ thể β :

$$\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n} = \frac{(\Delta' - \Delta_0)^2}{(z_\alpha + z_\beta)^2}$$

Khi 2 cỡ mẫu là bằng nhau, phương trình thu được:

$$m = n = \frac{(\sigma_1^2 + \sigma_2^2)(z_\alpha + z_\beta)^2}{(\Delta' - \Delta_0)^2}$$

Biểu thức cũng đúng với trường hợp α kiểm định một phía trái, và thay α bởi $\alpha/2$ ở trường hợp kiểm định 2 phía.

9.1.4 Kiểm định với mẫu cỡ lớn

Các giả định về phân phối chuẩn của tổng thể và các giá trị σ_1 và σ_2 may mắn không cần thiết khi cả hai cỡ mẫu đều đủ lớn. Trong trường hợp này, Định lý Giới hạn Trung tâm (Central Limit Theorem - CLT) đảm bảo rằng $\overline{X} - \overline{Y}$ xấp xỉ phân phối chuẩn bất kể phân phối của tổng thể như thế nào. Hơn nữa, sử dụng S_1^2 và S_2^2 thay thế cho σ_1^2 và σ_2^2 và trong Biểu thức (9.1) đưa ra một biến mà phân phối của nó xấp xỉ phân phối chuẩn chuẩn tắc:

$$Z = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}}$$

Một kết quả từ kiểm định thống kê với mẫu lớn bằng cách thay thế $\mu_1 - \mu_2$ bởi Δ_0, giá trị kỳ vọng của $\overline{X} - \overline{Y}$ khi H_0 đúng. Thống kê Z này sẽ có xấp xỉ phân phối chuẩn chuẩn tắc khi H_0 là đúng. Kiểm định với mức ý nghĩa mong muốn mà thu được bằng cách sử dụng một giá trị tới hạn z vẫn chính xác như trước đây.

Sử dụng giá trị kiểm định thống kê:

$$z = \frac{\overline{x} - \overline{y} - \Delta_0}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}}$$

cùng với các trường hợp miền bác bỏ bên phải, trái, và 2 phía dựa trên các giá trị tới hạn z cho kiểm định mẫu lớn mức ý nghĩa xấp xỉ α.
Những kiểm định này thường thích hợp nếu cả $m > 40$ và $n > 40$.
P-value cũng được tính chính xác như các bài kiểm định z trước đây.

Ví dụ 9.3 Tiêu thụ thức ăn nhanh có tác động gì đến sức khỏe và chế độ ăn uống không? (Từ tạp chí Nhi khoa Mỹ, 2004) báo cáo dữ liệu tóm tắt dưới đây về lượng calo hàng ngày cho một mẫu thiếu niên ít ăn thức ăn nhanh và một mẫu thanh thiếu niên khác thường ăn thức ăn nhanh.

Eat Fast Food	Sample Size	Sample Mean	Sample SD
No	663	2258	1519
Yes	413	2637	1138

Từ dữ liệu này có kết luận được lượng calo trung bình thực sự của thanh thiếu niên thường ăn thức ăn nhanh vượt quá 200 calo mỗi ngày so với mức tiêu thụ trung bình thực sự của những người không thường ăn thức ăn nhanh? Hãy kiểm định điều này với mức ý nghĩa 0.05.

9.1.5 Khoảng tin cậy cho $\mu_1 - \mu_2$

Khi cả hai phân phối tổng thể là phân phối chuẩn, việc chuẩn hóa $\overline{X} - \overline{Y}$ cho một biến ngẫu nhiên Z với phân phối chuẩn chuẩn tắc. Vì phần diện tích dưới đường cong của z nằm giữa $-z_{\alpha/2}$ và $z_{\alpha/2}$ là $1 - \alpha$, nên theo đó .

$$P\left(-z_{\alpha/2} < \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} < z_{\alpha/2}\right) = 1 - \alpha$$

Biến đổi các bất đẳng thức nằm trong ngoặc đơn để tách biệt $\mu_1 - \mu_2$ nhận được đẳng thức xác suất sau

$$P\left(\overline{X} - \overline{Y} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}} < \mu_1 - \mu_2 < \overline{X} - \overline{Y} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}\right) = 1 - \alpha$$

Điều này nghĩa là khoảng tin cậy $100(1-\alpha)$ % cho $\mu_1 - \mu_2$ có giới hạn dưới $\overline{x} - \overline{y} - z_{\alpha/2}.\sigma_{\overline{X}-\overline{Y}}$ và giới hạn trên $\overline{x} - \overline{y} + z_{\alpha/2}.\sigma_{\overline{X}-\overline{Y}}$, trong đó $\sigma_{\overline{X}-\overline{Y}}$ là biểu thức dạng căn bậc hai. Khoảng này là một trường hợp đặc biệt của công thức tổng quát $\hat{\theta} \pm z_{\alpha/2}.\sigma_{\hat{\theta}}$.

Nếu cả m và n đều lớn, Định lý giới hạn trung tâm chỉ ra rằng khoảng này có giá trị ngay cả khi không có giả định về các tổng thể chuẩn; trong trường hợp này, độ tin cậy xấp xỉ $100(1-\alpha)\%$. Hơn nữa, việc sử dụng sai số mẫu S_1^2 và S_2^2 trong biến chuẩn hóa Z thu được một khoảng hợp lệ trong đó s_1^2 và s_2^2 thay thế cho σ_1^2 và σ_2^2.

Nếu m và n đủ lớn, khoảng tin cậy cho $\mu_1 - \mu_2$ với mức độ tin cậy xấp xỉ $100(1-\alpha)\%$ là

$$\overline{x} - \overline{y} \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}$$

Trong đó dấu "–" cho giới hạn dưới và dấu "+" cho giới hạn trên của khoảng. Biên của khoảng tin cậy phía phải hoặc phía trái có thể được tính bằng cách giữ lại dấu hiệu (+ hoặc –) và thay thế $z_{\alpha/2}$ bởi z_α.

Nguyên tắc chung: kích thước mẫu được gọi là lớn khi $m > 40$ và $n > 40$.

Ví dụ 9.4 Một thí nghiệm được thực hiện để nghiên cứu các đặc điểm khác nhau của bu lông neo dẫn của 78 quan sát về cường độ (kip) của loại có đường kính 3/8 in, và 88 quan sát về cường độ của loại đường kính 1/2 in. Tóm tắt các giá trị thể hiện ở bảng bên dưới và biểu đồ hợp ở hình dưới. Kích cỡ mẫu, trung bình mẫu và độ lệch chuẩn mẫu phù hợp với các giá trị được đưa từ bài báo J. of Energy Engr., 1993. Hãy tìm khoảng tin cậy cho hiệu giữa cường độ thực sự của hai loại đường kính trên, với độ tin cậy 95 %

Variable	N	Mean	Median	TrMean	StDev	SEMean
diam 3/8	78	4.250	4.230	4.238	1.300	0.147

Variable	Min	Max	Q1	Q3
diam 3/8	1.634	7.327	3.389	5.075

Variable	N	Mean	Median	TrMean	StDev	SEMean
diam 1/2	88	7.140	7.113	7.150	1.680	0.179

Variable	Min	Max	Q1	Q3
diam 1/2	2.450	11.343	5.965	8.447

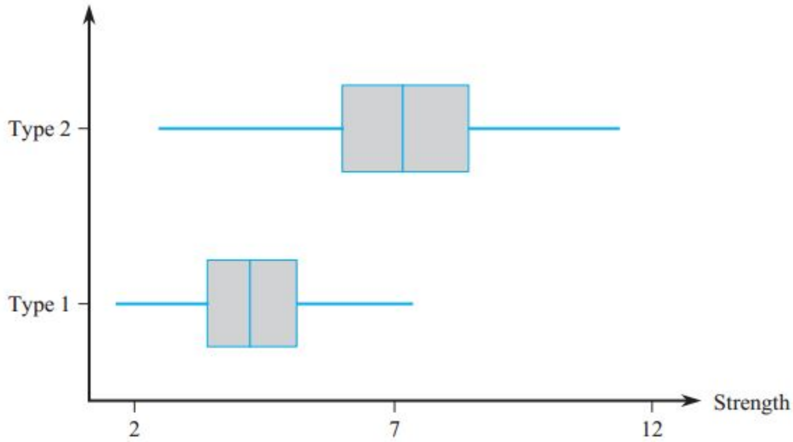


Figure 9.1 A comparative boxplot of the shear strength data

Nếu các phương sai σ_1^2 và σ_2^2 đều đã biết (hoặc biết giá trị xấp xỉ) và điều tra viên sử dụng các cỡ mẫu giống nhau, thì kích thước mẫu chung n để mà thu được một khoảng $100(1-\alpha)\%$ của chiều rộng w là

$$n = \frac{4z_{\alpha/2}^2(\sigma_1^2+\sigma_2^2)}{w^2}$$

(thông thường sẽ được làm tròn thành số nguyên).

9.2 Kiểm định t cho 2 mẫu và khoảng tin cậy

Thông thường các giá trị của phương sai tổng thể sẽ không được biết trước. Trong phần trước, chúng ta đã minh họa cho các cỡ mẫu lớn khi sử dụng một bài kiểm định z và khoảng tin cậy, trong đó các phương sai mẫu đã được sử dụng thay cho phương sai tổng thể. Thực tế, đối với các mẫu lớn, Định lý giới hạn trung tâm

cho phép chúng ta sử dụng các phương pháp này ngay cả khi hai tổng thể là không chuẩn.
Trong thực tế, mặc dù, thường xảy ra trường hợp ít nhất một mẫu có kích thước nhỏ và phương sai tổng thể là chưa biết. Nếu không dùng đến Định lý giới hạn trung tâm để loại bỏ, chúng ta có thể tiến hành bằng cách đưa ra các giả định cụ thể về phân phối tổng thể. Việc sử dụng các bước suy luận dựa theo những giả định này sẽ được giới hạn trong các tình huống mà các giả định ít nhất ở mức chấp nhận được. Ví dụ, chúng ta có thể giả định rằng cả hai phân phối tổng thể đều có phân phối Weibull hoặc cả hai đều là phân phối Poisson. Dù trong thực tế thì phân phối chuẩn thường là giả định hợp lý nhất.

Giả Thiết 9.2.1. *Cả 2 phân phối tổng thể là chuẩn, do đó $X_1, X_2, \ldots, X_m$ là mẫu ngẫu nhiên từ phân phối chuẩn và tương tự vậy cho $Y_1, \ldots, Y_n$ (với mẫu X và mẫu Y là độc lập với nhau). Khả năng hợp lý của các giả định này có thể được đánh giá bằng cách xây dựng một biểu đồ xác suất chuẩn của mẫu x_i và một biểu đồ khác của mẫu y_i.*

Công thức kiểm định thống kê và khoảng tin cậy dựa trên cùng một biến được chuẩn hóa đã được đề cập trong Phần 9.1, nhưng phân phối có liên quan bây giờ là t chứ không phải là z.

Định lý 9.2.2. *Khi các phân phối tổng thể là chuẩn, biến được chuẩn hóa*

$$T = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}} \tag{9.2}$$

sẽ xấp xỉ một phân phối t với bậc tư do ν được ước lượng từ dữ liệu bởi công thức

$$v = \frac{(\frac{s_1^2}{m} + \frac{s_2^2}{n})^2}{\frac{(s_1^2/m)^2}{m-1} + \frac{(s_2^2/n)}{n-1}} = \frac{[(se_1)^2 + (se_2)^2]^2}{\frac{(se_1)^4}{m-1} + \frac{(se_2)^4}{n-1}}$$

Trong đó

$$se_1 = \frac{s_1}{\sqrt{m}}, se_2 = \frac{s_2}{\sqrt{n}}$$

(ν được làm tròn xuống số nguyên gần nhất)

Biến đổi T trong biểu thức xác suất để tách riêng $\mu_1 - \mu_2$ cho khoảng tin cậy , trong khi một kiểm định thống kê thu được từ thay thế $\mu_1 - \mu_2$ bằng giá trị không Δ_0.

Khoảng tin cậy của $\mu_1 - \mu_2$ với kiểm định t cho 2 mẫu và độ tin cậy $100(1-\alpha)\%$ là

$$\overline{x} - \overline{y} \pm t_{\alpha/2,\nu}\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}$$

Khoảng tin cậy một phía có thể được tính như đã đề cập ở các phần trước.
Kiểm định t cho 2 mẫu để kiểm định giả thuyết $H_0 : \mu_1 - \mu_2 = \Delta_0$ như sau:

Giá trị kiểm định thống kê: $t = \frac{\overline{x} - \overline{y} - \Delta_0}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}}$

Các giả thuyết đối:	Vùng bác bỏ với kiểm định xấp xỉ mức ý nghĩa α
$H_a : \mu_1 - \mu_2 > \Delta_0$	$t \geq t_{\alpha,\nu}$ một phía phải
$H_a : \mu_1 - \mu_2 < \Delta_0$	$t \leq -t_{\alpha,\nu}$ một phía trái
$H_a : \mu_1 - \mu_2 \neq \Delta_0$	$t \geq t_{\alpha/2,\nu}$ hoặc $t \leq -t_{\alpha/2,\nu}$ hai phía.

Giá trị P có thể được tính như đã đề cập trong Phần 8.4 cho kiểm định t cho 1 mẫu.

Ví dụ 9.5 Lượng khoảng trống trong vải dệt ảnh hưởng đến sự thoải mái, dễ cháy và tính cách nhiệt. Tính thẩm thấu của vải là sự tiếp cận không gian trống đến dòng khí hoặc chất lỏng. Bài báo "Mối quan hệ giữa độ rỗng và tính thẩm thấu không khí của vải dệt thoi "(J. of Testing and Eval, 1997: 108-114) đưa ra thông tin tóm tắt về tính thấm khí ($cm^3/cm^2/$ giây) đối với một số loại vải khác nhau. Xem xét các dữ liệu sau đây trên hai loại vải:

Loại vải	**Cỡ mẫu**	**Trung bình mẫu**	**Độ lệch chuẩn mẫu**
cotton	10	51.71	0.79
triacetate	10	136.14	3.59

Giả sử phân phối độ rỗng cho cả hai loại vải là bình thường, chúng ta hãy tính khoảng tin cậy cho hiệu giữa độ rỗng trung bình thật sự cho vải bông và vải acetate, sử dụng độ tin cậy 95%.

Ví dụ 9.6 Việc xuống cấp của nhiều mạng lưới đường ống đô thị trên toàn quốc là một mối quan tâm ngày càng tăng. Một công nghệ đề xuất phục hồi đường ống sử dụng một lớp lót mềm dẻo luồn qua đường ống hiện có. Bài viết "Ảnh hưởng của hàn trên mật độ cao Polyethylene Liner "(J. của Vật liệu trong công trình dân dụng, 1996: 94–100) đã báo cáo các dữ liệu sau về độ bền kéo (psi) của mẫu lót khi có quá trình kết hợp được sử dụng và khi quá trình này không được sử dụng.

No fusion	2748	2700	2655	2822	2511			
	3149	3257	3213	3220	2753			
	$m = 10$	$\bar{x} = 2902.8$		$s_1 = 277.3$				
Fused	3027	3356	3359	3297	3125	2910	2889	2902
	$n = 8$	$\bar{y} = 3108.1$		$s_2 = 205.9$				

9.2.1 Quy trình t chung

Khác với quy trình kiểm định t cho 2 mẫu được mô tả trên là kết quả từ giả định rằng không chỉ hai phân phối tổng thể là chuẩn hóa mà còn có phương sai bằng nhau $\sigma_1^2 = \sigma_2^2$. Tức là, hai đường cong phân phối tổng thể được giả định là chuẩn, sự khác biệt duy nhất giữa chúng có thể là vị trí đối xứng của đường cong.
Đặt σ^2 là phương sai chung của tổng thể, thì chuẩn hóa của $\overline{X} - \overline{Y}$ là :

$$Z = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2}{m} + \frac{\sigma^2}{n}}} = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\sigma^2(\frac{1}{m} + \frac{1}{n})}}$$

sẽ có phân phối chuẩn hóa. Trước khi biến này có thể được sử dụng làm cơ sở cho việc kết luận về $\mu_1 - \mu_2$, phương sai chung phải được ước lượng từ dữ liệu mẫu. Một ước lượng của σ^2 là S_1^2, là phương sai của các m quan sát trong mẫu đầu tiên, và ước lượng khác là S_2^2 là phương sai của mẫu thứ hai. Theo một cách trực quan, một ước lượng tốt hơn so với từng phương sai mẫu cá nhân là kết hợp cả hai phương sai mẫu. Một ý tưởng đầu tiên có thể được sử dụng $(S_1^2 + S_2^2)/2$. Tuy nhiên, nếu $m > n$, thì mẫu đầu tiên chứa nhiều thông tin về σ^2 hơn mẫu thứ hai, và một nhận xét tương tự được áp dụng nếu $m < n$. Trọng số trung bình sau của hai phương sai mẫu, được gọi là **ước lượng chung của** σ^2 (hoặc kết hợp - pooled) , được điều chỉnh cho bất kỳ sự khác biệt nào giữa hai cỡ mẫu:

$$S_p^2 = \frac{m-1}{m+n-2}.S_1^2 + \frac{n-1}{m+n-2}.S_2^2$$

Mẫu đầu tiên có bậc tự do $m-1$ cho ước lượng của σ^2, và mẫu thứ hai có bậc tự do $n-1$, vì vậy bậc tự do tổng là $m+n-2$. Lý thuyết thống kê nói rằng nếu lấy S_p^2 thay thế σ^2 trong biểu thức của Z, biến chuẩn hóa sẽ có phân phối t với bậc tự do $m+n-2$. Trong cùng một cách mà các biến chuẩn hóa trước đây đã sử dụng làm cơ sở để tìm khoảng tin cậy và các bước kiểm định, biến t này ngay lập tức cũng cho khoảng tin cậy t chung để ước lượng $\mu_1-\mu_2$ và kiểm định t chung cho kiểm tra giả thuyết về hiệu hai trung bình.

Trong quá khứ, nhiều nhà thống kê đã đề xuất quy trình t chung này thay cho quy trình t hai mẫu. Ví dụ, kiểm định t chung có thể được rút ra từ nguyên tắc tỷ lệ hợp lý , trong khi đó kiểm định t hai mẫu không phải là một kiểm định tỷ lệ hợp lý. Hơn nữa, mức ý nghĩa cho bài kiểm định t chung là chính xác, trong khi nó chỉ là gần đúng với kiểm định t hai mẫu. Tuy nhiên, các nghiên cứu gần đây cho thấy mặc dù kiểm định t chung hiệu quả hơn so với kiểm định hai mẫu một chút (giá trị β nhỏ hơn với cùng α) khi mà $\sigma_1^2=\sigma_2^2$, thì bài kiểm định trước đây có thể dễ dẫn đến các kết luận sai nếu áp dụng khi các phương sai là khác nhau. Ý kiến tương tự áp dụng cho trường hợp của hai khoảng tin cậy. Do vậy, các quy trình t chung thực ra không vi phạm mạnh mẽ các giả định về phương sai bằng nhau.

Một gợi ý rằng có thể tiến hành kiểm định sơ bộ $H_0:\sigma_1^2=\sigma_2^2$ và sử dụng quy trình t chung nếu giả thuyết không không bị loại bỏ. Không may, kiểm định F thông dụng của các phương sai bằng nhau (Phần 9.5) khá nhạy với giả thiết phân phối tổng thể chuẩn - nhiều hơn so với quy trình t. Do đó khuyến khích nên sử dụng cách tiếp cận bảo toàn của quy trình t cho 2 mẫu, trừ khi có bằng chứng thực sự thuyết phục để thực hiện theo phương pháp khác, đặc biệt khi hai cỡ mẫu khác nhau.

9.2.2 Xác suất Sai lầm loại II

Xác định xác suất sai lầm loại II (hoặc tương đương, power $=1-\beta$) với kiểm định t cho 2 mẫu là phức tạp. Không đơn giản như cách sử dụng các đường cong β của Phụ lục Bảng A.17. Phiên bản gần đây nhất của Minitab (Phiên bản 16) sẽ tính toán power cho kiểm định t chung nhưng không phải cho kiểm định t cho hai mẫu.
Tuy nhiên, trang chủ của Bộ Thống kê (UCLA) (http://www.stat.ucla.edu) cho

phép truy cập vào một công cụ tính toán mạnh mẽ có thể làm điều này. Ví dụ, đặt $m = 10, n = 8, \sigma_1 = 300, \sigma_2 = 225$ (đây là các cỡ mẫu cho Ví dụ trên với độ lệch chuẩn mẫu nhỏ hơn một chút so với các giá trị của σ_1 và σ_2 trong ví dụ) và yêu cầu power của kiểm định 2 phía với mức ý nghĩa 0.05 cho giả thuyết $H_0 : \mu_1 - \mu_2 = 0$ khi $\mu_1 - \mu_2 = 100$, 250 và 500. Các giá trị kết quả của power lần lượt là 0.1089, 0.4609 và 0.9635 (tương ứng với $\beta = 0.89, 0.54$ và 0.04). Nói chung, β sẽ giảm khi kích thước mẫu tăng lên, lúc đó α tăng, và khi $\mu_1 - \mu_2$ di chuyển xa hơn từ 0. Phần mềm cũng sẽ tính các kích cỡ mẫu cần thiết để có được một giá trị cụ thể của power cho một giá trị cụ thể của $\mu_1 - \mu_2$.

9.3 Phân tích số liệu ghép đôi

Trong Phần 9.1 và 9.2, chúng ta đã xem xét kết luận về hiệu giữa hai giá trị trung bình μ_1 và μ_2 . Điều này được thực hiện bằng cách sử dụng các kết quả của mẫu ngẫu nhiên $X_1, X_2, .., X_m$ từ phân phối với trung bình μ_1 và mẫu hoàn toàn độc lập (với mẫu X) $Y_1, Y_2, \ldots, Y_n$ từ phân phối với giá trị trung bình μ_2. Đó là, hoặc m đối tượng được chọn từ tổng thể 1 và n đối tượng khác từ tổng thể 2, hoặc m cá thể thực nghiệm được điều trị theo cách này và một tập hợp khác của n cá thể được điều trị theo cách khác. Ngược lại, có một số tình huống thử nghiệm trong đó chỉ có một bộ n cá thể hoặc đối tượng thí nghiệm; và thực hiện hai quan sát trên một bộ và được một cặp các giá trị

Giả Thiết 9.3.1. *Giả định: Dữ liệu gồm n cặp được chọn độc lập* $(X_1, Y_1), (X_2, Y_2), \ldots$ (X_n, Y_n), *với* $E(X_i) = \mu_1$ *và* $E(Y_i) = \mu_2$. *Đặt* $D_1 = X_1 - Y_1, D_2 = X_2 - Y_2, \ldots D_n = X_n - Y_n$ *thì* D_i *là khác nhau từng cặp. Giả định mẫu* D_i *có phân phối chuẩn với giá trị trung bình* μ_D *và phương sai* σ_D^2 *(bởi vì mẫu* X_i *và mẫu* Y_i *có phân phối chuẩn). Chúng ta tiếp tục kết luận về hiệu của* $\mu_1 - \mu_2$. *Khoảng tin cậy cho kiểm định t hai mẫu và kiểm định thống kê thu được nhờ giả định mẫu độc lập áp dụng quy tắc* $V(\overline{X} - \overline{Y}) = V(\overline{X}) + V(\overline{Y})$. *Tuy nhiên với dữ liệu cặp, quan sát* X *và* Y *trong mỗi cặp thường không độc lập, cho nên* $\overline{X}$ *và* $\overline{Y}$ *không độc lập với nhau. Chúng ta vì vậy phải bỏ qua quy trình kiểm định t cho hai mẫu và tìm kiếm phương pháp thống kê thay thế.*

9.3.1 Kiểm định t cặp

Vì mỗi cặp khác nhau là độc lập, nên mẫu D_i cũng độc lập với nhau. Đặt $D = X - Y$, trong đó X và Y tương ứng là quan sát đầu tiên và thứ hai trong một cặp tùy ý. Vậy hiệu kỳ vọng là:

$$\mu_D = E(X - Y) = E(X) - E(Y) = \mu_1 - \mu_2$$

(quy tắc của các giá trị kỳ vọng được dùng ở đây là hợp lệ ngay cả khi X và Y phụ thuộc). Do đó bất kỳ giả thuyết nào về $\mu_1 - \mu_2$ có thể được diễn đạt như một giả thuyết về hiệu trung bình μ_D. Nhưng kể từ khi mẫu D_i tạo thành một mẫu ngẫu nhiên chuẩn (của hiệu) với trung bình μ_D, giả thuyết về μ_D có thể được kiểm định bằng cách sử dụng một kiểm định t một mẫu. Nghĩa là, để kiểm tra các giả thuyết về $\mu_1 - \mu_2$ khi dữ liệu là các cặp, ta tạo giá trị hiệu $D_1, D_2, \ldots, D_n$ và thực hiện một bài kiểm định t một mẫu (dựa trên bậc tự do $n-1$) trên các hiệu này.

Kiểm định t cặp.
Giả thuyết không: $H_0 : \mu_D = \Delta_0$ (trong đó $D = X - Y$ là hiệu giữa quan sát thứ nhất và quan sát thứ hai trong một cặp, và $\mu_d = \mu_1 - \mu_2$).
Giá trị thống kê: $t = \frac{\bar{d}-\Delta_0}{s_D/\sqrt{n}}$ (trong đó $\overline{D}$ và s_D là trung bình mẫu và độ lệch chuẩn tương ứng của mẫu d_i).

Các giả thuyết đối:	Miền bác bỏ với mức ý nghĩa α
$H_a : \mu_D > \Delta_0$	$t \geq t_{\alpha,n-1}$
$H_a : \mu_D < \Delta_0$	$t \leq -t_{\alpha,n-1}$
$H_a : \mu_D \neq \Delta_0$	$t \geq t_{\alpha/2,n-1}$ và $t \leq -t_{\alpha/2,n-1}$

Giá trị P-value được tính như các kiểm định t trước.

Ví dụ 9.7 Rối loạn cơ xương và vai khá phổ biến đối với nhân viên văn phòng, người thực hiện các tác vụ lặp lại bằng các đơn vị hiển thị trực quan. (bài báo của tạp chí Ergonomics, 1996: 1221–1230) báo cáo về một nghiên cứu để xác định liệu các điều kiện làm việc khác nhau có ảnh hưởng đến chuyển động của cánh tay hay không. Dữ liệu đi kèm được lấy từ một mẫu $n = 16$ đối tượng. Mỗi quan sát là lượng thời gian, được biểu thị bằng tỷ lệ tổng thời gian quan sát được, trong đó góc nâng của cánh tay là dưới 30^0. Hai số đo từ mỗi chủ đề đã được thu thập cách nhau 18 tháng. Trong thời gian này, điều kiện làm việc đã được thay đổi, và các đối tượng được phép tham gia vào một loạt các nhiệm vụ công việc khác nhau.

Dữ liệu có cho thấy thời gian trung bình thực trước và sau khi thay đổi nhiệm vụ có khác nhau với điều kiện khi mà góc nâng vẫn dưới 30^0 ?

Subject	1	2	3	4	5	6	7	8
Before	81	87	86	82	90	86	96	73
After	78	91	78	78	84	67	92	70
Difference	3	−4	8	4	6	19	4	3
Subject	9	10	11	12	13	14	15	16
Before	74	75	72	80	66	72	56	82
After	58	62	70	58	66	60	65	73
Difference	16	13	2	22	0	12	−9	9

Khi số lượng các cặp là lớn, giả định về phân phối chuẩn cho hiệu là không cần thiết. Định lý Giới hạn Trung tâm công nhận kết quả của kiểm định z.

9.3.2 Khoảng tin cậy của cặp.

Cùng cách mà khoảng tin cậy t cho giá trị trung bình tổng thể đơn μ được dựa trên biến t là $T = (\overline{X} - \mu)/(S/\sqrt{n})$, khoảng tin cậy t cho $\mu_D(= \mu_1 - \mu_2)$ dựa trên thực tế là

$$T = \frac{\overline{D} - \mu_D}{S_D/\sqrt{n}}$$

có phân phối t với bậc tự do $n - 1$. Biến đổi biến t như các phần trước đã làm với khoảng tin cậy, thu được khoảng tin cậy $100(1 - \alpha)$ % như sau:
Khoảng tin cậy cặp t cho μ_D là

$$\overline{d} \pm t_{\alpha/2,n-1}.s_D/\sqrt{n}$$

Khoảng tin cậy một phía thu được từ công thức trên bằng cách thay thế $t_{\alpha/2}$ bởi t_α.

Khi n nhỏ, tính hợp lệ của khoảng này yêu cầu phân phối của các hiệu phải xấp xỉ chuẩn. Với n lớn, Định lý Giới hạn Trung tâm đảm bảo cho kết quả của khoảng z là hợp lệ mà không có ràng buộc về phân phối của hiệu.

Ví dụ 9.8 Thêm hình ảnh y tế trên máy vi tính vào cơ sở dữ liệu hứa hẹn cung cấp tài nguyên tuyệt vời cho các bác sĩ. Tuy nhiên, có những phương pháp khác để thu thập thông tin như vậy, vì vậy vấn đề hiệu quả truy cập cần được điều tra. Bài báo " Hiệu quả so sánh của các thư viện hình ảnh thông thường và kỹ thuật số "(Tạp chí về phương tiện nghe nhìn trong y học, Mỹ, 2001: 8–15) đã báo cáo về một thử nghiệm trong đó 13 các chuyên gia y tế thông thạo máy tính đã được đặt giờ trong khi truy xuất hình ảnh từ thư viện các trang trình bày và trong khi truy xuất cùng một hình ảnh từ máy tính cơ sở dữ liệu với giao diện web.

Subject	1	2	3	4	5	6	7	8	9	10	11	12	13
Slide	30	35	40	25	20	30	35	62	40	51	25	42	33
Digital	25	16	15	15	10	20	7	16	15	13	11	19	19
Difference	5	19	25	10	10	10	28	46	25	38	14	23	14

Đặt μ_D là hiệu trung bình thật sự của thời gian truy xuất slide (giây) và thời gian truy xuất kỹ thuật số. Sử dụng khoảng tin cậy t cặp để ước lượng μ_D với phân phối hiệu này xấp xỉ phân phối chuẩn.

9.3.3 Dữ liệu cặp và quy trình kiểm định t cho hai mẫu

Xem xét kiểm định t cho hai mẫu trên dữ liệu cặp. Các tử số của hai kiểm định thống kê là đồng nhất, vì $\overline{d} = \Sigma d_i/n = [\Sigma(x_i - y_i)]/n = (\Sigma x_i)/n - (\Sigma y_i)/n = \overline{x} - \overline{y}$. Sự khác nhau giữa số liệu thống kê là hoàn toàn do mẫu số . Mỗi kiểm định thống kê thu được bằng cách chuẩn hóa $\overline{X} - \overline{Y}(= \overline{D})$. Nhưng với sự hiện diện của sự phụ thuộc, việc chuẩn hóa hai mẫu t là không chính xác. Để thấy điều này, nhớ lại Phần 5.5 rằng

$$V(X \pm Y) = V(X) + V(Y) \pm 2Cov(X, Y)$$

Tương quan giữa X và Y là:

$$\rho = Corr(X, Y) = Cov(X, Y)/[\sqrt{V(X)}.\sqrt{V(Y)}]$$

Theo đó:

$$V(X - Y) = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$$

Áp dụng vào $\overline{X} - \overline{Y}$ nhận được

$$\boxed{V(\overline{X} - \overline{Y}) = V(\overline{D}) = V(\tfrac{1}{n}\Sigma D_i) = \tfrac{V(D_i)}{n} = \tfrac{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}{n}}$$

Kiểm định t cho hai mẫu dựa trên giả định về tính độc lập, trong trường hợp $\rho = 0$. Tuy nhiên, trong nhiều thí nghiệm cặp, sẽ có sự phụ thuộc dương mạnh mẽ giữa X và Y (X lớn liên quan đến Y lớn), do đó ρ sẽ dương và phương sai của $\overline{X} - \overline{Y}$ sẽ nhỏ hơn $\sigma_1^2/n + \sigma_2^2/n$. Do vậy, *bất cứ khi nào có sự phụ thuộc dương trong cặp, mẫu số cho thống kê t cặp phải nhỏ hơn so với t của kiểm định mẫu độc lập.* Thông thường t hai mẫu sẽ gần với số không hơn so với t cặp, giảm bớt đáng kể ý nghĩa của dữ liệu.
Tương tự như vậy, khi dữ liệu dạng cặp, khoảng tin cậy t cặp thường sẽ hẹp hơn khoảng tin cậy (không chính xác) của t cho hai mẫu . Điều này là vì thông thường sự biến thiên trong giá trị hiệu ít hơn so với biến thiên của giá trị x và y.

9.3.4 Thử nghiệm ghép cặp và không ghép cặp

Trong ví dụ của chúng ta, kết quả dữ liệu cặp từ hai quan sát về cùng chủ đề hoặc đối tượng thực nghiệm . Ngay cả khi không thể thực hiện được, dữ liệu cặp với sự phụ thuộc trong cặp có thể thu được bằng cách kết hợp các cá nhân hoặc đối tượng trên một hoặc nhiều đặc tính ảnh hưởng đến sự hồi đáp. Ví dụ, trong một thử nghiệm y học để so sánh hiệu quả của hai loại thuốc giảm huyết áp, ngân sách của thí nghiệm có thể cho phép điều trị 20 bệnh nhân. Nếu 10 bệnh nhân được lựa chọn ngẫu nhiên để điều trị với thuốc thứ nhất và 10 lựa chọn độc lập khác để điều trị với thuốc thứ hai, kết quả thử nghiệm mẫu độc lập .
Tuy nhiên, người thử nghiệm, biết rằng huyết áp chịu ảnh hưởng bởi tuổi và cân nặng, có thể quyết định tạo ra cặp bệnh nhân để trong mỗi cặp, có 10 cặp, tuổi và cân nặng xấp xỉ nhau (mặc dù có thể có sự khác biệt lớn giữa các cặp). Sau đó, mỗi loại thuốc sẽ được trao cho bệnh nhân khác nhau trong mỗi cặp cho tổng cộng 10 quan sát trên mỗi loại thuốc.

Nếu không có sự kết hợp này (hoặc "chặn"), một loại thuốc có thể có hiệu quả tốt hơn chỉ vì bệnh nhân trong một mẫu bị nhẹ hơn và trẻ hơn và do đó dễ bị giảm huyết áp hơn những bệnh nhân nặng hơn và lớn tuổi hơn trong mẫu thứ hai. Tuy nhiên, có sự đánh đổi khi thực hiện ghép cặp – số bậc tự do ít hơn khi phân tích cặp - vì vậy chúng ta phải xem xét khi nào nên ưu tiên loại thử nghiệm nào hơn.

Không có câu trả lời đơn giản và chính xác cho câu hỏi này, nhưng có một số hướng dẫn hữu ích. Nếu chúng ta có sự lựa chọn giữa hai kiểm định t hợp lệ (và được thực hiện ở cùng mức ý nghĩa α), chúng ta nên ưu tiên kiểm định có số bậc tự do lớn hơn. Lý do cho điều này là bậc tự do lớn hơn có nghĩa là β nhỏ hơn cho bất kỳ giá trị thay thế cố định của tham số hoặc các tham số . Nghĩa là, so với xác suất sai lầm loại I cố định, xác suất sai lầm loại II được giảm đi khi tăng số bậc tự do.

Tuy nhiên, nếu các đơn vị thực nghiệm không đồng nhất trong dấu hiệu hồi đáp của chúng thì sẽ rất khó phát hiện ra sự khác biệt nhỏ có ý nghĩa giữa các biện pháp điều trị. Nếu có tương quan dương cao trong các đơn vị thực nghiệm hoặc các đối tượng, phương sai của $\overline{D} = \overline{X} - \overline{Y}$ sẽ nhỏ hơn nhiều so với phương sai không cặp. Do phương sai giảm đi này, sẽ dễ dàng hơn để phát hiện ra sự khác biệt với các mẫu cặp so với các mẫu độc lập. Các ưu và khuyết điểm của ghép cặp bây giờ có thể được tóm tắt như sau.

1. Nếu có sự không đồng nhất lớn giữa các đơn vị thực nghiệm và mối tương quan lớn trong các đơn vị thực nghiệm (ρ dương và lớn), thì sự mất mát về bậc tự do sẽ được bù đắp bằng độ chính xác tăng liên quan đến việc ghép cặp, do đó, một kiểm định ghép cặp là tốt hơn một kiểm định mẫu độc lập.
2. Nếu các đơn vị thực nghiệm tương đối đồng nhất và mối tương quan giữa các cặp không lớn, độ chính xác do ghép cặp sẽ bị ảnh hưởng nhiều hơn bởi sự giảm bậc tự do, do đó nên sử dụng kiểm định mẫu độc lập.

Tất nhiên, các giá trị của σ_1^2, σ_2^2, và ρ thường không được biết chính xác, do đó, điều tra viên sẽ phải thực hiện các phỏng đoán có cơ sở về tình huống 1 hoặc 2 nhận được. Nói chung, nếu số lượng quan sát có thể thu được là lớn thì sự mất mát về bậc tự do (ví dụ, từ 40 đến 20) sẽ không nghiêm trọng; nhưng nếu số quan sát này nhỏ, thì sự mất mát này (từ 16 đến 8) vì ghép cặp có thể là nghiêm trọng nếu không được bù đắp bằng tăng độ chính xác. Các cân nhắc tương tự được áp dụng khi chọn giữa hai loại thực nghiệm để ước tính $\mu_1 - \mu_2$ với một khoảng tin cậy.

9.4 Các kết luận liên quan đến hiệu hai tỷ lệ

Sau khi trình bày các phương pháp so sánh trung bình của hai tổng thể khác nhau, chúng ta chuyển sự chú ý đến việc so sánh tỉ lệ của tổng thể. Cho rằng một cá nhân hoặc đối tượng được đánh giá là thành công "S" nếu anh/chị/đối tượng có được một số đặc trưng (ví dụ: người đã tốt nghiệp đại học, sở hữu chiếc tủ lạnh có chức năng tạo đá viên,...). Đặt

p_1= thành phần đạt tiêu chí S trong tổng thể #1

p_2= thành phần đạt tiêu chí S trong tổng thể #2

Ngoài ra, $p_1(p_2)$ có thể được coi là xác suất một cá nhân (hoặc đối tượng) ngẫu nhiên được chọn từ tổng thể đầu tiên (thứ hai) và thành công.
Giả sử một mẫu có kích thước m được chọn từ tổng thể thứ nhất và một mẫu độc lập kích thước n được chọn từ tổng thể thứ hai. Đặt X biểu thị số S trong mẫu đầu tiên và Y là số S trong mẫu thứ hai. Sự độc lập của hai mẫu suy ra rằng X và Y là độc lập. Nếu hai kích cỡ mẫu nhỏ hơn nhiều so với kích cỡ tổng thể tương ứng, X và Y có thể được coi là có phân phối nhị thức. Ước lượng tự nhiên cho $p_1 - p_2$ (hiệu về tỉ lệ tổng thể), là tương ứng với hiệu tỷ lệ mẫu $X/m - Y/n$.

Mệnh đề 9.4.1. *Đặt $\hat{p_1} = X/m$ và $\hat{p_2} = Y/n$, trong đó $X \sim Bin(m, p_1), Y \sim Bin(n, p_2)$, với Xvà Y là các biến độc lập. Vậy*

$$E(\hat{p_1} - \hat{p_2}) = p_1 - p_2$$

vì $\hat{p_1} - \hat{p_2}$ là ước lượng không chệch của $p_1 - p_2$, và

$$V(\hat{p_1} - \hat{p_2}) = \frac{p_1q_1}{m} + \frac{p_2q_2}{n} \tag{9.3}$$

với $q_i = 1 - p_i$

Đầu tiên chúng ta sẽ tập trung vào các tình huống trong đó cả m và n đều lớn. Sau đó vì mỗi $\hat{p_1}$ và $\hat{p_2}$ đều có phân phối xấp xỉ chuẩn, ước lượng của $\hat{p_1} - \hat{p_2}$ cũng có phân phối xấp xỉ chuẩn. Chuẩn hóa $\hat{p_1} - \hat{p_2}$ nhận được biến Z có phân phối xấp xỉ chuẩn:

$$Z = \frac{\hat{p_1} - \hat{p_2} - (p_1 - p_2)}{\sqrt{\frac{p_1q_1}{m} + \frac{p_2q_2}{n}}}$$

9.4.1 Quy trình kiểm định mẫu lớn

Giả thuyết không mà một nhà điều tra hay gặp nhất sẽ có dạng $H_o : p_1 - p_2 = \Delta_0$. Mặc dù với trung bình tổng thể thì trường hợp $\Delta_0 \neq 0$ là không có gì khó khăn, thì với tỉ lệ của tổng thể $\Delta_0 = 0$ và $\Delta_0 \neq 0$ phải được xem xét riêng. Vì đa số các vấn đề thực tế của loại này thường liên quan $\Delta_0 = 0$ (như giả thuyết không $p_1 = p_2$), chúng ta sẽ tập trung vào trường hợp này. Khi $H_o : p_1 - p_2 = 0$ là đúng, đặt p kí hiệu giá trị chung của p_1 và p_2 (và tương tự cho q). Thì biến chuẩn hóa

$$Z = \frac{\hat{p_1} - \hat{p_2} - 0}{\sqrt{pq(\frac{1}{m} + \frac{1}{n})}} \tag{9.4}$$

sẽ có xấp xỉ phân phối chuẩn chuẩn tắc khi H_0 là đúng. Tuy nhiên, Z này không thể dùng như một kiểm định thống kê vì giá trị của p là chưa biết , H_0 chỉ khẳng định ở đây là có một giá trị chung của p, nhưng không nói giá trị đó là bao nhiêu. Một kiểm định thống kê là kết quả việc thay thế p và q trong (9.4) bằng các ước lượng thích hợp.

Giả định rằng $p_1 = p_2 = p$, thay vì các mẫu riêng biệt kích thước m và n từ hai tổng thể khác nhau (hai phân phối nhị phân khác nhau), chúng ta thực sự có một mẫu đơn cỡ $m + n$ từ một quần thể với tỷ lệ p. Tổng số cá thể trong mẫu kết hợp này có đặc trưng là $X + Y$. Ước lượng tự nhiên của p là:

$$\hat{p} = \frac{X + Y}{m + n} = \frac{m}{m + n}.\hat{p_1} + \frac{n}{m + n}.\hat{p_2} \tag{9.5}$$

Biểu thức thứ hai cho $\hat{p}$ thấy rằng nó thực sự là một trung bình trọng số của các ước lượng $\hat{p_1}$ và $\hat{p_2}$ thu được từ hai mẫu. Sử dụng $\hat{p}$ và $\hat{q} = 1 - \hat{p}$ thay cho của p và q trong (9.4) đưa ra một kiểm định thống kê có phân phối chuẩn chuẩn tắc khi H_0 là đúng.

Ví dụ 9.9 Bài báo "Sử dụng Aspirin và sự sống còn sau khi chẩn đoán ung thư đại trực tràng" (J. Amer. Med. PGS., 2009: 649–658) báo cáo rằng 549 người tham gia nghiên cứu thường dùng aspirin sau khi được chẩn đoán mắc bệnh ung thư đại trực tràng, đã có 81 người tử vong, trong khi đó 730 người được chẩn đoán tương tự thì không sử dụng aspirin có 141 người bệnh chết. Liệu rằng dữ liệu này có cho kết luận rằng việc sử dụng aspirin thường xuyên sẽ giảm tỷ lệ tử vong sau khi được chuẩn đoán ung thư đại trực tràng? Hãy kiểm định các giả thuyết với mức ý nghĩa là 0.05.

Giả thuyết không: $H_0 : p_1 - p_2 = 0$
Giá trị kiểm định thống kê (mẫu lớn) : $z = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\hat{p}\hat{q}(\frac{1}{m} + \frac{1}{n})}}$

Các giả thuyết đối:	Miền bác bỏ với kiểm định xấp xỉ mức α
$H_a : p_1 - p_2 > 0$	$z \geq z_\alpha$
$H_a : p_1 - p_2 < 0$	$z \leq -z_\alpha$
$H_a : p_1 - p_2 \neq 0$	$z \geq z_{\alpha/2}$ hoặc $z \geq -z_{\alpha/2}$

Giá trị P-value được tính như các kiểm định z trước. Kiểm định có thể dùng được khi mà $m\hat{p_1}, m\hat{q_1}, n\hat{p_2}$ và $n\hat{q_2}$ đều tối thiểu là 10.

9.4.2 Xác suất sai lầm loại II và cỡ mẫu

Ở đây việc xác định β là một chút rườm rà hơn so với trước đây ở các kiểm định với mẫu lớn khác. Lý do là mẫu số của Z là ước lượng độ lệch chuẩn của $\hat{p} - \hat{p_2}$, giả sử rằng $p_1 = p_2 = p$. Khi H_0 là sai, $\hat{p_1} - \hat{p_2}$ phải được chuẩn hóa lại bằng cách dùng

$$\sigma_{\hat{p_1} - \hat{p_2}} = \sqrt{\frac{p_1 q_1}{m} + \frac{p_2 q_2}{n}} \tag{9.6}$$

Biểu thức của σ ngụ ý rằng β không phải là một hàm của chỉ $p_1 - p_2$, vì vậy chúng ta biểu thị nó bằng $\beta(p_1, p_2)$.

Các giả thuyết đối	$\beta(p_1, p_2)$.
$H_a : p_1 - p_2 > 0$	$\Phi[\frac{z_\alpha \sqrt{\overline{p}.\overline{q}(\frac{1}{m} + \frac{1}{n})} - (p_1 - p_2)}{\sigma}]$
$H_a : p_1 - p_2 < 0$	1- $\Phi[\frac{-z_\alpha \sqrt{\overline{p}.\overline{q}(\frac{1}{m} + \frac{1}{n})} - (p_1 - p_2)}{\sigma}]$
$H_a : p_1 - p_2 \neq 0$	$\Phi[\frac{z_{\alpha/2} \sqrt{\overline{p}.\overline{q}(\frac{1}{m} + \frac{1}{n})} - (p_1 - p_2)}{\sigma}] - \Phi[\frac{-z_{\alpha/2} \sqrt{\overline{p}.\overline{q}(\frac{1}{m} + \frac{1}{n})} - (p_1 - p_2)}{\sigma}]$

với $\overline{p} = (mp_1 + np_2)/(m + n), \overline{q} = (mq_1 + nq_2)/(m + n)$ và σ được cho bởi công thức (9.6) .

Ngoài ra, đối với cụ thể p_1, p_2 mà $p_1 - p_2 = d$, có thể xác định cỡ mẫu cần có được $\beta(p_1, p_2) = \beta$. Ví dụ: đối với kiểm định bên phải, chúng ta tương đương $-z_\beta$ với argument của $\Phi(.)$ (tức là những gì nằm trong ngoặc đơn) được nói trong phần đóng khung trên. Nếu $m = n$, có một biểu thức đơn giản cho giá trị chung.

Trường hợp $m = n$, kiểm định với mức ý nghĩa α có xác suất sai lầm loại II β ở các giá trị thay thế p_1, p_2 với $p_1 - p_2 = d$ khi

$$n = \frac{[z_\alpha\sqrt{(p_1+p_2)(q_1+q_2)/2}+z_\beta\sqrt{p_1q_1+p_2q_2}]^2}{d^2} \quad (9.7)$$

với kiểm định bên phải hoặc bên trái, thay thế $\alpha/2$ vào α với kiểm định 2 phía.

Ví dụ 9.10 Một trong những ứng dụng thống kê thực sự ấn tượng là sự kết nối với thiết kế thí nghiệm vắc-xin bại liệt Salk 1954 và phân tích dữ liệu kết quả. Một phần của thí nghiệm tập trung vào hiệu quả của vắc-xin trong việc chống bệnh bại liệt. Bởi vì người ta nghĩ rằng không có một nhóm trẻ em đối chứng, sẽ có không có cơ sở để đánh giá vắc-xin. Người ta quyết định tiêm vắc xin cho 1 nhóm, còn 1 nhóm tiêm giả dược(không thể phân biệt bằng mắt giữa vắc xin và giả dược) đối với nhóm đối chứng. Vì lý do đạo đức và cũng bởi vì khi sự nhận biết có vắc xin có thể ảnh hưởng đến điều trị và chẩn đoán, thí nghiệm được tiến hành trong cách thức mù đôi (double-blind). Tức là, cá nhân sẽ không biết mình được tiêm vắc xin hay giả dược (các mẫu được mã hóa bằng số). (Hãy nhớ rằng: tại thời điểm đó mọi người chưachắc rằng vắc xin có tác dụng tốt)
Cho p_1 và p_2 là xác suất một đứa trẻ bị bại liệt cho nhóm đối chứng và nhóm điều trị. Tiêu chuẩn kiểm định $H_0 : p_1-p_2 = 0$ và $H_a : p_1-p_2 > 0$ (nghĩa là trẻ được tiêm vắc xin ít bị bại liệt hơn trẻ không tiêm). Giả sử giá trị thật của p_1 là 0.0003 , vắc xin sẽ là một bước phát triển lớn nếu mức độ này giảm một nửa hay $p_2 = 0.00015$. Sử dụng mức ý nghĩa $\alpha = 0.05$, thì $\beta = 0.1$ khi $p_1 = 0.0003$ và $p_2 = 0.00015$. Giả sử cỡ mẫu bằng nhau, tìm n?

9.4.3 Khoảng tin cậy cho mẫu lớn

Giống như trung bình, nhiều vấn đề hai mẫu liên quan đến vấn đề của so sánh thông qua kiểm định giả thuyết, nhưng đôi khi ước lượng khoảng cho $p_1 - p_2$ là thích hợp. Cả $\hat{p_1} = X/m$ và $\hat{p_2} = Y/n$ đều có phân phối xấp xỉ chuẩn khi m và n lớn. Nếu chúng ta xác định θ với $p_1 - p_2$, thì $\hat{\theta} = \hat{p_1} - \hat{p_2}$ đáp ứng các điều kiện cần thiết để có được một khoảng tin cậy của mẫu lớn. Cụ thể, ước lượng độ lệch tiêu chuẩn $\hat{\theta}$ là $\sqrt{(\hat{p_1}\hat{q_1}/m) + (\hat{p_2}\hat{q_2}/n)}$. Khoảng tin cậy tổng quát $100(1-\alpha)\%$ là $\hat{\theta} \pm z_{\alpha/2}.\hat{\sigma}_{\hat{\theta}}$ có dạng như sau.

Khoảng tin cậy $p_1 - p_2$ với độ tin cậy xấp xỉ $100(1-\alpha)\%$ là

$$\hat{p^1} - \hat{p^2} \pm z_{\alpha/2}\sqrt{\frac{\hat{p^1}.\hat{q_1}}{m} + \frac{\hat{p_2}\hat{q_2}}{n}}$$

Khoảng này có thể dùng được khi mà $m\hat{p_1}, m\hat{q_1}, n\hat{p_2}$ và $n\hat{q_2}$ đều tối thiểu là 10.

Chú ý rằng độ lệch chuẩn được ước lượng của $\hat{p_1} - \hat{p_2}$ (dạng căn thức) thì khác với khi kiểm định giả thuyết khi $\Delta_0 = 0$.
Các nghiên cứu gần đây cho thấy độ tin cậy thực sự của khoảng tin cậy truyền thống đôi khi chệch khỏi mức đa thức (mức mà ta nhận được khi sử dụng giá trị đặc trưng z, ví dụ 95% khi $z_{\alpha/2} = 1.96$). Cải tiến đề xuất là thêm một thành công và một thất bại cho hai mẫu và sau đó thay thế các giá trị mẫu $\hat{p}$ và mẫu $\hat{q}$ trong công thức nói trên bằng mẫu $\tilde{p}$ và mẫu $\tilde{q}$, trong đó $\tilde{p_1} = (x+1)/(m+2)$,v.v... Khoảng hiệu chỉnh này cũng có thể được sử dụng khi kích thước mẫu là khá nhỏ.

Ví dụ 9.11 Các tác giả của bài báo "Liệu pháp xạ trị bổ trợ và hóa trị liệu ở phụ nữ tiền mãn kinh NodePositive với ung thư vú" (New Engl. J. of Med., 1997: 956–962) báo cáo về kết quả của một thử nghiệm được thiết kế để so sánh bệnh nhân ung thư dùng hóa trị và bệnh nhân dùng kết hợp hóa trị và xạ trị. Trong số 154 cá nhân nhận được điều trị hóa trị, 76 người sống sót ít nhất 15 năm, trong khi 98 người sống lâu trong số 164 bệnh nhân được điều trị cả hóa trị và xạ trị. Với độ tin cậy 99 % tìm khoảng tin cây cho hiệu giữa hai tỉ lệ trên?

9.4.4 Suy luận với mẫu nhỏ

Thỉnh thoảng suy luận liên quan $p_1 - p_2$ có thể phải dựa trên trường hợp có ít nhất một mẫu nhỏ. Các phương pháp thích hợp cho các tình huống như vậy không đơn giản như đối với các mẫu lớn và có nhiều tranh cãi giữa các nhà thống kê về các bước tiến hành. Một kiểm định thường xuyên được sử dụng, được gọi là kiểm định Fisher-Irwin, được dựa trên phân phối siêu bội.

9.5 Các kết luận liên quan đến hai phương sai tổng thể

Đôi khi cần đến phương pháp để so sánh hai phương sai tổng thể (hoặc độ lệch chuẩn), mặc dù các vấn đề này ít gặp hơn so với trung bình hoặc tỷ lệ. Đối với trường hợp trong đó các tổng thể đang được điều tra là chuẩn, các bước dựa trên một họ mới của phân phối xác suất mới.

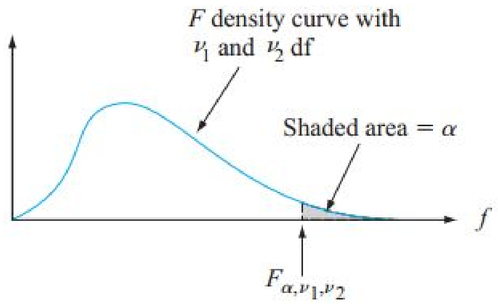
9.5.1 Phân phối F

Phân phối xác suất F có hai tham số, ký hiệu là ν_1 và ν_2. Tham số ν_1 được gọi là *số bậc tự do của tử* , và ν_2 là *số bậc tự do của mẫu* ; ở đây ν_1 và ν_2 là số nguyên dương. Một biến ngẫu nhiên có phân phối F không thể giả sử là một giá trị âm. Vì hàm mật độ là phức tạp và sẽ không được sử dụng rõ ràng, chúng ta bỏ qua công thức. Có một mối quan hệ quan trọng giữa một biến F và biến chi bình phương. Nếu X_1 và X_2 là các biến ngẫu nhiên có phân phối chi bình phương độc lập với bậc tự do ν_1 và ν_2 tương ứng, thì biến ngẫu nhiên

$$F = \frac{X_1/v_1}{X_2/v_2}$$

(tỉ số của 2 biến chi bình phương chia cho bậc tự do tương ứng), có thể được cho thấy để có phân phối F.

Hình dưới minh họa đồ thị của một hàm mật độ F điển hình. Tương tự với ký hiệu $t_{\alpha,v}$ và $\chi^2_{\alpha,v}$, chúng ta sử dụng F_{α,v_1,v_2} cho giá trị trên trục hoành mà α nằm trong vùng diện tích dưới đường cong mật độ F với bậc tự do ν_1 và ν_2 ở miền bên phải. Đường cong mật độ không đối xứng, vì vậy có vẻ như các giá trị tới hạn của phía trái và phải đều phải được lập bảng. Tuy nhiên điều này không cần thiết vì thực tế là $F_{1-\alpha,v_1,v_2} = 1/F_{\alpha,v_2,v_1}$.



Phụ lục Bảng A.9 cho F_{α,v_1,v_2} , và $\alpha = 0.10, 0.05, 0.01$ và 0.001, và các giá trị khác của ν_1 (trong các cột khác nhau của bảng) và ν_2 (trong các nhóm hàng khác nhau của bảng). Ví dụ, $F_{0.05,6,10} = 3.22$ và $F_{0.05,10,6} = 4.06$. Giá trị tới hạn $F_{0.95,6,10}$ mà chiếm 0.95 diện tích miền phải nó (và như vậy 0.05 ở bên trái) dưới đường cong F với $v_1 = 6$ và $v_2 = 10$, là $F_{0.95,6,10} = 1/F_{0.05,6,10} = 1/4.06 = 0.246$.

9.5.2 Kiểm định F cho các phương sai bằng nhau

Các bước kiểm định cho giả thuyết liên quan tỉ lệ σ_1^2/σ_2^2 dựa trên kết quả sau:

Định lý 9.5.1. *Đặt $X_1, \ldots, X_m$ là một mẫu ngẫu nhiên có phân phối chuẩn với phương sai σ_1^2, đặt $Y_1, \ldots, Y_n$ là một mẫu ngẫu nhiên khác (độc lập với mẫu X_i) có phân phối chuẩn với phương sai σ_2^2 , và đặt S_1^2 và S_2^2 biểu thị hai phương sai mẫu. Thì biến ngẫu nhiên*

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

có phân phối F với $v_1 = m - 1$ và $v_2 = n - 1$.

Định lý này là kết quả từ việc kết hợp (9.8) với thực tế là các biến $(m-1)S_1^2/\sigma_1^2$ và $(n-1)S_2^2/\sigma_2^2$ mỗi biến đều có phân phối chi bình phương với bậc tự do $m-1$ và $n-1$ tương ứng (xem Phần 7.4). Bởi vì F bao gồm một tỷ lệ chứ không phải là một hiệu , kiểm định thống kê là tỷ số của hai phương sai mẫu. Yêu cầu $\sigma_1^2 = \sigma_2^2$ bị loại bỏ nếu tỷ lệ khác nhau quá nhiều so với 1.

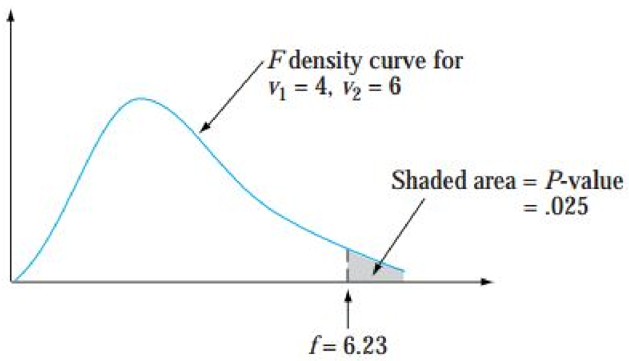
Khi giá trị tới hạn được cho trong bảng chỉ cho $\alpha = 0.10, 0.05, 0.01$ và 0.001, kiểm định 2 phía chỉ có thể thực hiện ở mức $0.20, 0.10, 0.02, 0.002$. Các giá trị tới hạn F khác có thể thu được bằng phần mềm thống kê.

Giả thuyết không: $H_0 : \sigma_1^2 = \sigma_2^2$
Giá trị kiểm định thống kê: $f = s_2^1/s_2^2$

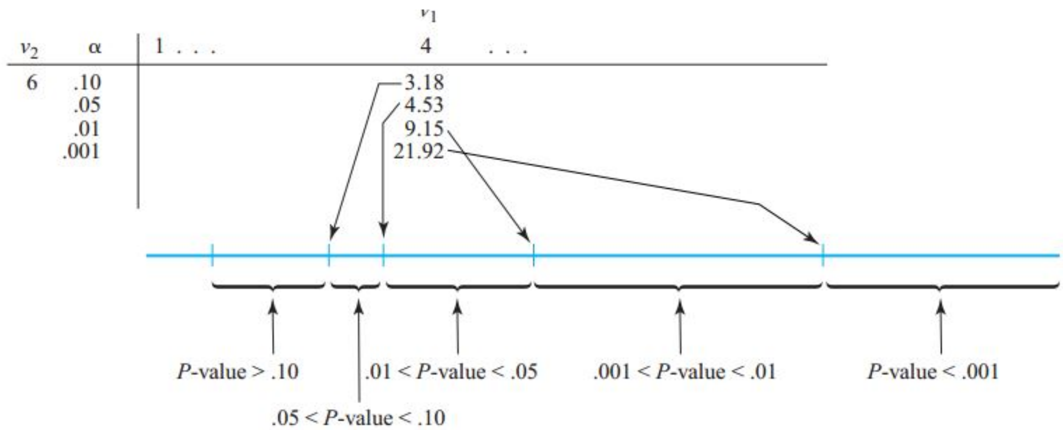
Các giả thuyết đối:	Miền bác bỏ cho a với mức ý nghĩa α
$H_a : \sigma_1^2 > \sigma_2^2$	$f \geq F_{\alpha,m-1,n-1}$
$H_a : \sigma_1^2 < \sigma_2^2$	$f \leq F_{\alpha,m-1,n-1}$
$H_a : \sigma_1^2 \neq \sigma_2^2$	$f \geq F_{\alpha/2,m-1,n-1}$ hay $f \leq F_{\alpha/2,m-1,n-1}$

9.5.3 P-value cho kiểm định F

Nhắc lại về $P - value$ cho một kiểm định t bên phải trên là phần diện tích dưới đường cong t tương ứng (phần với bậc tự do phù hợp) đến bên phải của giá trị t được tính toán. Cùng một cách, giá trị P cho một kiểm định F miền phải là phần diện tích dưới đường cong F với bậc tự do của tử số và mẫu số thích hợp khi kiểm định bên phải của f đã được tính. Hình 9.9 minh họa điều này với một bài kiểm định dựa trên $v_1 = 4$ và $v_2 = 6$.



Lập bảng của phần diện tích miền phải đường cong F phức tạp hơn nhiều so với đường cong t bởi vì liên quan đến hai bậc tự do . Đối với mỗi sự kết hợp của v_1 và v_2, bảng F của chúng ta chỉ đưa ra bốn giá trị tới hạn mà diện tích là $0.10, 0.05, 0.01$, và 0.001. Hình dưới cho thấy những gì có thể nói về P-value phụ thuộc vào nơi f rơi vào tương ứng với bốn giá trị tới hạn.



Ví dụ, kiểm định với $v_1 = 4$ và $v_2 = 6$,
$f = 5.70 \Rightarrow 0.01 <$P-value$< 0.05$
$f = 2.16 \Rightarrow$P-value> 0.10
$f = 25.03 \Rightarrow$P-value< 0.001
Chỉ khi f bằng một giá trị đã lập bảng thì chúng ta mới có được một giá trị P chính xác (ví dụ, nếu $f = 4.53$, thì $P - value = 0.05$). Một khi chúng ta biết rằng $0.01 < P - value < 0.05$, H_0 sẽ bị loại bỏ ở mức ý nghĩa 0.05 nhưng không ở mức 0.01. Khi $P - value < 0.001$, H_0 bị từ chối ở bất kỳ mức ý nghĩa hợp lý nào.

Kiểm định F được thảo luận trong các chương tiếp theo sẽ là phía phải. Tuy nhiên, nếu một kiểm định F bên trái là thích hợp, thì các giá trị tới hạn của phía trái được tìm như các mô tả đã đề cập để các giới hạn trên P-value có thể được thiết lập. Trong trường hợp kiểm định hai phía, các giới hạn từ một bài kiểm định một phía được nhân với 2. Ví dụ, nếu $f = 5.82$ khi $v_1 = 4$ và $v_2 = 6$, thì vì 5.82 rơi vào giữa giá trị 0.05 và 0.01 và $2(0.01) < P - value < 2(0.05)$, nên $0.02 < P - value < 0.10$. H_0 sẽ bị từ chối nếu $\alpha = 0.10$ nhưng không phải với $\alpha = 0.01$. Trong trường hợp này, không thể nói vì từ trong bảng của chúng ta về kết luận nào là phù hợp khi $\alpha = 0.05$ (vì chúng ta không biết $P - value$ có nhỏ hơn hay lớn hơn). Tuy nhiên, phần mềm thống kê cho thấy diện tích bên phải của 5.82 dưới đường cong F là 0.029, do vậy giá P-value là 0.058 và giả thuyết không vì thế không nên bị bác bỏ ở mức 0.05 (0.058 là α nhỏ nhất mà H_0 có thể bị từ chối và giá trị α mà chúng ta đã chọn nhỏ hơn giá trị này). Tất nhiên các phần mềm thống kê sẽ cung cấp $P - value$ chính xác cho bất kỳ kiểm định F nào.

9.5.4 Khoảng tin cậy cho σ_1/σ_2

Khoảng tin cậy cho σ_1^2/σ_2^2 là dựa trên thay thế F trong biểu thức xác suất

$$P(F_{1-\alpha/2,v_1,v_2} < F < F_{\alpha/2,v_1,v_2}) = 1 - \alpha$$

bằng biến F ở công thức trên và biến đổi các bất đẳng thức để tách riêng σ_1^2/σ_2^2 . Một khoảng cho σ_1/σ_2 là kết quả từ lấy căn bậc hai của mỗi giới hạn. Các phần chi tiết được chừa lại như một bài tập.

Chương 10

PHÂN TÍCH PHƯƠNG SAI

Giới thiệu

10.1 Một nhân tố ANOVA

10.2 So sánh trong ANOVA

10.3 Nhiều hơn trên một nhân tố ANOVA

Chương 11

PHÂN TÍCH PHƯƠNG SAI NHIỀU NHÂN TỐ

Giới thiệu

11.1 Hai nhân tố ANOVA với $K_{ij} = 1$

11.2 Hai nhân tố ANOVA với $K_{ij} > 1$

11.3 Ba nhân tố ANOVA

11.4 NHân tố 2^p

Chương 12

TƯƠNG QUAN VÀ HỒI QUI TUYẾN TÍNH

Giới thiệu

12.1 Mô hình hồi qui tuyến tính

12.2 Ước lượng tham số mô hình

12.3 Suy luận về hệ số dốc β_1

12.4 Suy luận về sự liên quan giữa $\mu_{Y,x}$ và giá trị dự đoán của Y

12.5 Hệ số tương quan

TỪ KHÓA

Tài liệu tham khảo