

TCVN

TIÊU CHUẨN QUỐC GIA

TCVN 14253:2024

ISO 11132:2021

Xuất bản lần 1

**PHÂN TÍCH CẢM QUAN – PHƯƠNG PHÁP LUẬN –
HƯỚNG DẪN ĐO LƯỜNG NĂNG LỰC CỦA HỘI ĐỒNG
CẢM QUAN MÔ TẢ ĐỊNH LƯỢNG**

*Sensory analysis – Methodology – Guidelines for the measurement
of the performance of a quantitative descriptive sensory panel*

HÀ NỘI – 2024

Lời nói đầu

TCVN 14253:2024 hoàn toàn tương đương với ISO 11132:2021;

TCVN 14253:2024 do Ban kỹ thuật tiêu chuẩn quốc gia TCVN/TC/F13 *Phương pháp phân tích và lấy mẫu* biên soạn, Viện Tiêu chuẩn Chất lượng Việt Nam đề nghị, Ủy ban Tiêu chuẩn Đo lường Chất lượng Quốc gia thẩm định, Bộ Khoa học và Công nghệ công bố.

Lời giới thiệu

Hội đồng đánh giá có thể được sử dụng làm công cụ để xác định các thuộc tính cảm quan của sản phẩm và đánh giá mức độ của các thuộc tính cảm quan.

Năng lực là thước đo khả năng của một hội đồng hoặc một người đánh giá trong việc đưa ra các đánh giá thuộc tính hợp lệ và đáng tin cậy trên các sản phẩm được đánh giá. Hội đồng có thể được đánh giá tại một thời điểm nhất định, thường là sau một thời gian huấn luyện (thẩm tra) hoặc được theo dõi theo thời gian (giám sát). Năng lực bao gồm khả năng của một hội đồng trong việc phát hiện, nhận diện và đo lường một thuộc tính, sử dụng các thuộc tính theo cách tương tự với các hội đồng khác hoặc giữa các thành viên trong một hội đồng, phân biệt các tác nhân kích thích, sử dụng thang đo đúng cách, lặp lại kết quả của chính họ và tái lập kết quả trong so sánh với các hội đồng hoặc người đánh giá khác.

Việc đo lường năng lực cho phép chủ tịch hội đồng cải thiện kết quả đầu ra của hội đồng và người đánh giá, nhận diện các vấn đề và nhu cầu huấn luyện lại hoặc nhận diện những người đánh giá không thực hiện đủ tốt để tiếp tục tham gia đánh giá.

Phân tích cảm quan – Phương pháp luận – Hướng dẫn đo lường năng lực của hội đồng cảm quan mô tả định lượng

Sensory analysis – Methodology – Guidelines for the measurement of the performance of a quantitative descriptive sensory panel

1 Phạm vi áp dụng

Tiêu chuẩn này đưa ra hướng dẫn để đánh giá năng lực tổng thể của hội đồng mô tả định lượng và năng lực của từng thành viên hội đồng.

Tiêu chuẩn này có thể áp dụng để thẩm tra việc huấn luyện hội đồng hoặc từng thành viên hội đồng, cũng như để giám sát năng lực của hội đồng đã thành lập.

Tiêu chuẩn này không áp dụng cho năng lực của hội đồng đối với các phương pháp mô tả khi không ghi lại điểm riêng của từng thành viên hội đồng, khi không có danh mục riêng các thuộc tính cho tất cả những người đánh giá hoặc khi đo lường tính ưu thế thay vì đo cường độ. Do đó, năng lực của các hội đồng mô tả sử dụng các phương pháp như profile đồng thuận, profile lựa chọn tự do, profile nhanh và tính ưu thế theo thời gian của cảm giác (TDS) không thuộc phạm vi của tiêu chuẩn này.

Các phương pháp được quy định trong tiêu chuẩn này là để theo dõi và đánh giá khả năng của hội đồng và thành viên hội đồng trong việc phân biệt các sản phẩm, sự đồng thuận giữa các thành viên trong cùng một hội đồng và độ lặp lại của các thành viên trong việc chấm điểm cường độ.

Độ tái lập, bao gồm cả việc so sánh giữa các hội đồng và so sánh trong cùng một hội đồng gồm nhiều cuộc đánh giá được thực hiện trong các điều kiện khác nhau (nghĩa là tách biệt về mặt thời gian) không thuộc phạm vi của tiêu chuẩn này.

Chủ tịch hội đồng có thể sử dụng toàn bộ hoặc chỉ một số phương pháp quy định trong tiêu chuẩn này để đánh giá thường xuyên năng lực của các hội đồng hoặc từng thành viên trong hội đồng. Các phương pháp được liệt kê trong tiêu chuẩn này là chưa đầy đủ và có thể sử dụng các phương pháp thích hợp khác.

2 Tài liệu viện dẫn

Các tài liệu viện dẫn sau rất cần thiết cho việc áp dụng tiêu chuẩn này. Đối với các tài liệu viện dẫn ghi năm công bố thì áp dụng phiên bản được nêu. Đối với các tài liệu viện dẫn không ghi năm công bố thì áp dụng phiên bản mới nhất, bao gồm cả các sửa đổi, bổ sung (nếu có).

TCVN 11182 (ISO 5492), *Phân tích cảm quan – Thuật ngữ và định nghĩa*

3 Thuật ngữ và định nghĩa

Trong tiêu chuẩn này, áp dụng các thuật ngữ và định nghĩa nêu trong TCVN 11182 (ISO 5492) cùng với các thuật ngữ và định nghĩa sau:

3.1

Sự đồng thuận (agreement)

Khả năng của các hội đồng hoặc người đánh giá khác nhau cùng thể hiện sự khác biệt của sản phẩm khi chấm điểm đối với một thuộc tính nhất định cho cùng một bộ sản phẩm

3.2

Sự lệch hướng của hội đồng (panel drift)

Hiện tượng trong đó hội đồng thay đổi độ nhạy hoặc trở nên dễ bị chệch theo thời gian và do đó làm thay đổi vị trí của một thuộc tính trên thang đo nơi thuộc tính đó được xếp loại so với sản phẩm tham chiếu

3.3

Năng lực/hiệu quả hoạt động (performance)

Khả năng của một hội đồng hoặc một người đánh giá để đưa ra những đánh giá đáng tin cậy và hợp lệ về các kích thích và thuộc tính của kích thích

3.4

Xác nhận giá trị sử dụng/thẩm tra (validation)

Quá trình xác định rằng một hội đồng hoặc một người đánh giá có thể đáp ứng các tiêu chí *năng lực* (3.3) cụ thể

3.5

Phiên đánh giá/buổi đánh giá (session)

Khoảng thời gian mà sản phẩm được đánh giá

CHÚ THÍCH 1: Trong một phiên đánh giá riêng lẻ, một hoặc nhiều sản phẩm có thể được đánh giá bởi một hoặc nhiều người đánh giá. Đối với người đánh giá, dù thực hiện một mình hay là thành viên hội đồng, các phiên đánh giá đều được phân tách theo thời gian.

[NGUỒN: 4.63 của ISO 5492:2008/Amd.1:2016]

3.6

Lặp lại (replicate)

Sự xuất hiện của một điều kiện cụ thể trong thiết kế thực nghiệm

CHÚ THÍCH 1: Thuật ngữ này thường ngụ ý rằng sự xuất hiện là một trong nhiều sự kiện cùng loại, nhưng nó có thể đề cập đến một sự xuất hiện đơn lẻ. Khi điều kiện được thực hiện hai lần, cách diễn đạt là "hai lần lặp lại" v.v...

CHÚ THÍCH 2: Thuật ngữ "lặp lại" hoặc "phiên đánh giá lặp lại" là rõ ràng hơn để xác định nhiều lần xuất hiện của một điều kiện.

CHÚ THÍCH 3: "Phiên đánh giá lặp lại" là *phiên đánh giá* (3.5) trong đó người đánh giá, sản phẩm, điều kiện thử và mục tiêu là như nhau.

3.7

Độ chệch của người đánh giá (assessor bias)

Xu hướng của người đánh giá chấm điểm số luôn cao hơn hoặc thấp hơn so với điểm thực khi đã biết điểm thực hoặc so với điểm trung bình của hội đồng khi chưa biết điểm thực

[NGUỒN: 1.40 của ISO 5492:2008/Amd.1:2016]

3.8

Độ chệch do thứ tự bố trí (order bias)

Độ chệch phát sinh từ vị trí tương đối về không gian hoặc thời gian của sản phẩm so với nhóm sản phẩm được đánh giá

CHÚ THÍCH 1: Thuật ngữ này bao gồm cả "độ chệch vị trí" và "độ chệch tuần tự".

[NGUỒN: 1.42 của ISO 5492:2008/Amd.1:2016]

3.9

Độ lặp lại (repeatability)

Sự đồng thuận (3.1) trong việc đánh giá các sản phẩm giống nhau trong cùng điều kiện thử nghiệm bởi cùng một người đánh giá hoặc cùng một hội đồng đánh giá

CHÚ THÍCH 1: Độ lặp lại có thể được đo lường trong một *phiên đánh giá* (3.5) hoặc qua nhiều phiên riêng biệt rõ ràng, với điều kiện là các đánh giá *lặp lại* (3.6) được tiến hành trong các điều kiện thử nghiệm có thể được coi là như nhau. Nếu việc đánh giá lặp lại được tiến hành trong các phiên đánh giá/buổi đánh giá riêng biệt rõ ràng thì các phiên đánh giá thường chỉ cách nhau vài ngày. Trong trường hợp này, sự khác biệt giữa độ lặp lại và *độ tái lập* (3.10) trong ngắn hạn là rất nhỏ và liên quan đến các điều kiện thử nghiệm có được coi là như nhau hay không.

[NGUỒN: 1.45 của ISO 5492:2008/Amd.1:2016, đã sửa đổi – bổ sung Chú thích 1.]

3.10

Độ tái lập (reproducibility)

Sự đồng thuận (3.1) trong việc đánh giá cùng một sản phẩm trong các điều kiện thử nghiệm khác nhau hoặc bởi những người đánh giá hoặc hội đồng khác nhau

TCVN 14253:2024

CHÚ THÍCH 1: Độ tái lập có thể được đo lường bằng các hình thức sau đây:

- độ tái lập ngắn hạn của một hội đồng (hoặc người đánh giá) được đo giữa hai hoặc nhiều *phiên đánh giá* (3.5) cách nhau vài ngày;
- độ tái lập trung hạn hoặc dài hạn của một hội đồng (hoặc người đánh giá), được đo giữa các phiên đánh giá cách nhau vài tháng;
- độ tái lập giữa các hội đồng khác nhau, trong cùng một phòng thử nghiệm hoặc trong các phòng thử nghiệm khác nhau.

[NGUỒN: 1.46 của ISO 5492:2008/Amd.1:2016]

4 Nguyên tắc

4.1 Hai cách tiếp cận khả thi

4.1.1 Yêu cầu chung

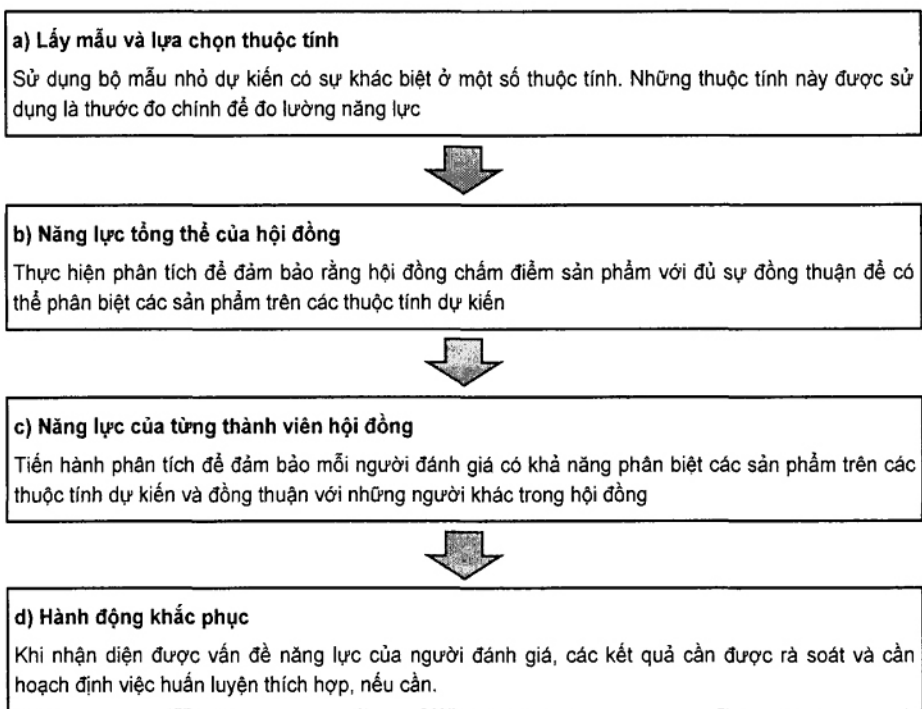
Tiêu chuẩn này liên quan đến các hội đồng cảm quan được sử dụng để đánh giá độ lớn của một hoặc nhiều thuộc tính cảm quan nhằm tạo ra các mô tả định lượng hoặc profile của sản phẩm [xem TCVN 12749 (ISO 13299)]. Các phương pháp khác thích hợp để đo lường năng lực của các hội đồng được sử dụng để thử nghiệm sự khác biệt.

Năng lực của hội đồng cảm quan định lượng có thể được đánh giá từ các phiên đánh giá được tiến hành riêng nhằm đạt được năng lực (được gọi là “quy trình chuyên biệt”) bằng cách sử dụng các đánh giá đã có sẵn (được gọi là “giám sát liên tục”).

4.1.2 Đo lường năng lực thông qua quy trình chuyên biệt

Một quy trình chuyên biệt là phương pháp được lựa chọn để cấp chứng chỉ cho mỗi người đánh giá và cho các mục đích xác nhận khác. Đối với việc gia hạn chứng nhận, quy trình chuyên biệt này phải được lặp lại định kỳ, nếu cần. Hình 1 là lưu đồ cho quy trình này.

Cách tiếp cận này thường có thể được sử dụng ở cuối giai đoạn huấn luyện của hội đồng để đảm bảo rằng hội đồng và từng thành viên hội đồng đã đạt được mức năng lực mong muốn và có thể được coi là người đánh giá cảm quan đã được huấn luyện hoặc chuyên gia thử cảm quan (tùy thuộc vào tiêu chí năng lực).



Hình 1 – Các bước xử lý để đo lường năng lực thông qua quy trình chuyên biệt

4.1.3 Giám sát liên tục thông qua việc lập profile sản phẩm thường quy

Một cách tiếp cận khác bao gồm giám sát dữ liệu profile đã được thu thập. Để xem xét dữ liệu profile liên tục do một hội đồng tạo ra, có thể phù hợp khi sử dụng dữ liệu có nguồn gốc từ các thử nghiệm lập profile khá khác nhau bằng cách sử dụng các loại sản phẩm, số sản phẩm khác nhau, v.v... Quy trình này giống như quy trình được nêu trong Hình 1. Tuy nhiên, vì không có sự khác biệt được xác định trước, các thuộc tính mà sản phẩm có sự phân biệt đáng kể bởi toàn hội đồng cho một profile nhất định được sử dụng làm thước đo chính để kiểm tra năng lực của từng thành viên hội đồng cảm quan. Các thuộc tính không tạo ra sự khác biệt đáng kể thì không được sử dụng làm thước đo đáng tin cậy để kiểm tra sự nhất quán vì thiếu sự đồng thuận trong nội bộ và giữa những người đánh giá cảm quan có thể có nghĩa là các sản phẩm rất giống nhau về các đặc tính đó.

Trong trường hợp này, trong một khoảng thời gian nhất định, cần phải kiểm tra một bộ sản phẩm khác biệt hơn để đảm bảo hội đồng thực sự có khả năng làm nổi bật sự khác biệt về các đặc điểm này.

4.2 Các chỉ số về năng lực của hội đồng hoặc từng thành viên hội đồng

Đối với một cuộc đánh giá, có thể xác định các chỉ số sau:

- khả năng phân biệt của hội đồng, được đo bằng khả năng của hội đồng cho thấy sự khác biệt đáng

TCVN 14253:2024

kể giữa các sản phẩm;

- khả năng phân biệt của người đánh giá, được đo bằng khả năng của người đánh giá cho thấy sự khác biệt đáng kể giữa các sản phẩm;
- sự đồng thuận của người đánh giá, được đo bằng mức độ phù hợp giữa điểm sản phẩm trung bình của người đánh giá và điểm trung bình của hội đồng;
- sự đồng thuận của hội đồng, được đo bằng mức độ phù hợp giữa điểm sản phẩm trung bình giữa những người đánh giá.

Đối với đánh giá lặp lại:

- độ lặp lại của người đánh giá, được đo bằng mức độ đồng nhất giữa các lần đánh giá lặp lại đối với cùng một sản phẩm;
- độ lặp lại của hội đồng, được đo bằng mức độ đồng nhất trung bình giữa các lần đánh giá lặp lại của cùng một sản phẩm đối với mỗi người đánh giá.

4.3 Phân tích thống kê

Cách tiếp cận thống nhất và duy nhất để phân tích thống kê các kết quả được mô tả trong tiêu chuẩn này. Tuy nhiên, một vài chỉ số về năng lực của hội đồng có thể được đánh giá bằng nhiều thước đo. Ví dụ: bình phương của sai số trung bình và độ lệch chuẩn của sai số (SD) (căn bậc hai của nó) đều thể hiện sự biến thiên trong việc đánh giá một sản phẩm. Các thước đo được sử dụng phải là những thước đo thông dụng trong lĩnh vực áp dụng.

Các thước đo liên quan khác về sự đồng thuận giữa những người đánh giá trong việc sử dụng thang đo cho một thuộc tính là sự tương tác giữa người đánh giá và sản phẩm cũng như hệ số tương quan giữa điểm của người đánh giá và điểm trung bình của hội đồng. Người đánh giá có thể không có độ chệch nhưng có thể sử dụng thang đo theo một cách khác. Mỗi tương quan gần bằng 1, độ dốc hồi quy gần bằng 1 và giao điểm hồi quy gần bằng 0 cho thấy sự đồng thuận cao giữa người đánh giá và những người còn lại trong hội đồng.

Khi mỗi người đánh giá thực hiện đánh giá một số lượng nhỏ mẫu (ít hơn sáu), hệ số tương quan cần được diễn giải thận trọng vì hệ số tương quan có thể cao (đến 0,7) chỉ do ngẫu nhiên.

5 Điều kiện tiên quyết

5.1 Điều kiện thực nghiệm

Cơ sở thử nghiệm nên phù hợp với TCVN 12390 (ISO 8589).

5.2 Trình độ chuyên môn của người đánh giá

Những người đánh giá được lựa chọn/sàng lọc trong hội đồng cần có trình độ chuyên môn và kinh nghiệm phù hợp với TCVN 12389 (ISO 8586) hoặc tiêu chuẩn cao hơn.

6 Đo lường năng lực thông qua quy trình chuyên biệt

6.1 Lựa chọn mẫu và thuộc tính

Tại mỗi nghiên cứu chuyên biệt, hội đồng đánh giá cần được giới thiệu một bộ mẫu tương tự với các mẫu mà hội đồng sẽ thực hiện khi đánh giá sản phẩm và dự kiến sẽ có sự khác biệt có ý nghĩa thống kê giữa ít nhất một cặp mẫu đối với các thuộc tính liên quan.

Để đảm bảo rằng tất cả các khía cạnh chính của sản phẩm đều được kiểm tra, cần đưa vào thử nghiệm một tập hợp các thuộc tính đa dạng đầy đủ.

Các thuộc tính liên quan này được sử dụng làm thước đo chính để đánh giá năng lực của hội đồng. Bộ mẫu cần bao gồm các mẫu lặp lại. Mỗi mẫu cần có cùng số mẫu lặp lại. Các mẫu lặp lại có thể được đánh giá trong một phiên đánh giá hoặc qua hai phiên trở lên. Số lượng người đánh giá, số mẫu và số lần lặp lại phụ thuộc vào sản phẩm, các thuộc tính cảm quan cần đánh giá và mục đích của quy trình đánh giá. Ví dụ, có thể sử dụng hai hoặc ba lần lặp lại của ba hoặc bốn mẫu. Cần thận trọng hạn chế số lần đánh giá cần thiết trong một phiên đánh giá để tránh mệt mỏi về cảm giác. Các thuộc tính của mẫu phải tương tự trong dải giá trị mà hội đồng thực hiện khi đánh giá sản phẩm.

6.2 Thiết kế thực nghiệm

6.2.1 Yêu cầu chung

Một số loại thiết kế thực nghiệm có thể được sử dụng trong quy trình chuyên biệt, tùy thuộc vào mục tiêu quan trọng nhất cần trả lời.

6.2.2 Thiết kế khối ngẫu nhiên

Có thể sử dụng thiết kế thực nghiệm khối ngẫu nhiên, trong đó người đánh giá là các "khối". Thiết kế này phù hợp khi dự kiến không có ảnh hưởng (hiệu ứng) từ mẫu này sang mẫu tiếp theo. Nếu ngược lại thì nên sử dụng thiết kế cân bằng thay thế (xem 6.2.3).

6.2.3 Thiết kế cân bằng và ngẫu nhiên

Nếu dự kiến có ảnh hưởng từ mẫu này sang mẫu tiếp theo thì thiết kế thử nghiệm phù hợp là hình vuông Williams Latin ^[12]. Bảng 1 thể hiện thiết kế hình vuông Williams Latin với bốn người đánh giá và bốn mẫu.

Bảng 1 – Thiết kế hình vuông Williams Latin cho bốn người đánh giá và bốn mẫu

Người đánh giá	Phiên đánh giá	Thứ tự			
		1	2	3	4
1	1	A	B	C	D
2	1	B	D	A	C
3	1	C	A	D	B
4	1	D	C	B	A
1	2	B	D	A	C
2	2	C	A	D	B
3	2	D	C	B	A
4	2	A	B	C	D

Trong thiết kế này, mỗi người đánh giá sẽ lấy mẫu bốn sản phẩm theo thứ tự bố trí khác nhau trong một phiên đánh giá nhất định và mỗi sản phẩm cụ thể sẽ được lấy bởi một người đánh giá. Ví dụ: trong phiên 1, người đánh giá 1 lấy A và B, người đánh giá 2 lấy C, người đánh giá 3 lấy D và người đánh giá 4 không lấy mẫu nào.

Đối với mỗi lần lặp lại đánh giá sản phẩm, nên sử dụng thứ tự bố trí sản phẩm khác nhau cho mỗi người đánh giá để giảm ảnh hưởng do thứ tự bố trí và ảnh hưởng chuyển tiếp.

Nếu có bội số của bốn người đánh giá thì có thể lặp lại thiết kế tương tự cho mỗi bộ bốn người.

Cũng có thể lựa chọn thiết kế thứ tự sản phẩm ngẫu nhiên, tức là tác động ngẫu nhiên từng sản phẩm đến từng vị trí trong mỗi phiên đánh giá.

Ưu điểm của cách tiếp cận này là giảm thiểu ảnh hưởng chuyển tiếp ở cấp độ hội đồng và do đó có được ước tính tốt hơn về trung bình sản phẩm ở cấp độ hội đồng để đánh giá năng lực. Tuy nhiên, nếu thứ tự bố trí sản phẩm có tác động thì sự đồng thuận giữa những người đánh giá sẽ bị ảnh hưởng vì mỗi người đánh giá sẽ không trải nghiệm cùng một thứ tự bố trí sản phẩm. Để so sánh những người đánh giá trong cùng một mục đích, có thể sử dụng cùng một thứ tự bố trí sản phẩm cho tất cả những người đánh giá cảm quan (xem 6.2.4).

6.2.4 Thiết kế với cùng thứ tự bố trí

Để tập trung vào năng lực của từng người đánh giá và trong nỗ lực so sánh những người đánh giá trong những điều kiện giống nhau nhất, một thiết kế thay thế được đề xuất, theo đó tất cả những người đánh giá đều thử các sản phẩm theo cùng một thứ tự bố trí, xem Bảng 2.

Bảng 2 – Thiết kế cùng thứ tự bố trí cho bốn người đánh giá và bốn mẫu

Người đánh giá	Phiên đánh giá	Thứ tự			
		1	2	3	4
1	1	A	B	C	D
2	1	A	B	C	D
3	1	A	B	C	D
4	1	A	B	C	D
1	2	A	B	C	D
2	2	A	B	C	D
3	2	A	B	C	D
4	2	A	B	C	D

Trong trường hợp này, người đánh giá không đánh giá bản thân sản phẩm mà đánh giá sản phẩm ở một vị trí nhất định (hiệu ứng sản phẩm và vị trí bị lẫn lộn). Điều này sẽ dẫn đến ước tính sai lệch về tác động của sản phẩm (bị chệch bởi ảnh hưởng do thứ tự bố trí), nhưng không dẫn đến ước tính sai lệch về tác động của người đánh giá và sự tương tác giữa sản phẩm*người đánh giá.

6.3 Phân tích thống kê

Bảng 3 minh họa một cách lập bảng và tóm tắt kết quả. Một số phần mềm máy tính có thể yêu cầu cách tổ chức dữ liệu khác, ví dụ: điền các mẫu theo các cột và người đánh giá theo các hàng.

Việc đo lường năng lực của hội đồng trên toàn bộ hội đồng và từng thành viên hội đồng, ngoài độ chệch, dữ liệu cần được phân tích bằng phân tích phương sai (ANOVA) [7].

Chi tiết về các phép tính cơ bản không được trình bày trong tiêu chuẩn này vì việc phân tích thường được thực hiện bằng phần mềm.

Bảng 3 – Kết quả đánh giá trên một thuộc tính

Mẫu	Người đánh giá								Trung bình
	1		...	j		...	n _q		
	Điểm	Trung bình		Điểm	Trung bình		Điểm	Trung bình	
1	Y ₁₁₁ Y ₁₁₂ ... Y _{11n_r}	$\bar{Y}_{11.}$		Y _{1j1} Y _{1j2} ... Y _{1jn_r}	$\bar{Y}_{1j.}$				$\bar{Y}_{1..}$
...									
i	Y _{i11} Y _{i12} ... Y _{i1n_r}	$\bar{Y}_{i1.}$		Y _{ij1} Y _{ij2} ... Y _{ijn_r}	$\bar{Y}_{ij.}$				$\bar{Y}_{i..}$
...									
n _p									$\bar{Y}_{n_p..}$
Trung bình	$\bar{Y}_{.1.}$			$\bar{Y}_{.j.}$			$\bar{Y}_{.n_q.}$		$\bar{Y}_{...}$
Trong bảng này giả định rằng: n _p = số lượng mẫu (i = 1, 2 ... n _p); n _q = số lượng người đánh giá (j = 1, 2 ... n _q); n _r = số lần lặp trên mỗi mẫu (k = 1, 2 ... n _r).									

Dữ liệu của mỗi người đánh giá được phân tích bằng ANOVA một chiều khi các đánh giá lặp lại được tiến hành trong một phiên duy nhất (xem Bảng 4).

Nếu các đánh giá lặp lại được thực hiện trong các phiên riêng biệt thì có thể áp dụng mô hình ANOVA một chiều (tức là hiệu ứng mẫu) hay mô hình ANOVA hai chiều (tức là hiệu ứng phiên đánh giá và mẫu) (xem Bảng 4 và Bảng 5), tùy thuộc vào thông lệ chuẩn của nhà nghiên cứu/phòng thử nghiệm.

Bảng 4 – ANOVA đối với từng thành viên hội đồng cho một thuộc tính

Nguồn biến thiên	Bậc tự do	Tổng bình phương	Bình phương trung bình	Tỷ lệ F
Giữa các mẫu	$v_1 = n_p - 1$	S_1	$MS_1 = S_1/v_1$	$F = MS_1/MS_2$
Sai số	$v_2 = n_p(n_r - 1)$	S_2	$MS_2 = S_2/v_2$	
Tổng cộng	$v_3 = n_p n_r - 1$	S_3		
n_p = số lượng mẫu; n_r = số lần lặp trên mỗi mẫu.				

Bảng 5 – ANOVA đối với từng thành viên hội đồng cho một thuộc tính có tính đến hiệu ứng phiên

Nguồn biến thiên	Bậc tự do	Tổng bình phương	Bình phương trung bình	Tỷ lệ F
Giữa các mẫu	$v_1 = n_p - 1$	S_1	$MS_1 = S_1/v_1$	$F = MS_1/MS_5$
Giữa các phiên đánh giá	$v_4 = n_s - 1$	S_4	$MS_4 = S_4/v_4$	$F = MS_4/MS_5$
Sai số	$v_5 = n_p(n_r - 1) - (n_s - 1)$	S_5	$MS_5 = S_5/v_5$	
Tổng cộng	$v_3 = n_p n_r - 1$	S_3		
n_p = số lượng mẫu; n_s = số phiên đánh giá; n_r = số lần lặp trên mỗi mẫu.				

Bộ dữ liệu hoàn chỉnh được phân tích theo ANOVA khối ngẫu nhiên (xem Bảng 6).

Hiệu ứng người đánh giá có thể được coi là cố định hoặc ngẫu nhiên^[6]. Để đo lường năng lực, thông thường người ta đặt hiệu ứng của người đánh giá là cố định vì trọng tâm là năng lực của những người đánh giá cụ thể tham gia vào nghiên cứu. Tuy nhiên, để dự đoán tốt hơn năng lực trong các điều kiện đánh giá thực tế, cũng có thể chọn yếu tố người đánh giá là ngẫu nhiên (xem chú thích a của Bảng 6 và Bảng 7).

Bộ dữ liệu hoàn chỉnh được phân tích bằng ANOVA hai chiều khi các đánh giá lặp lại được tiến hành trong một phiên duy nhất (xem Bảng 6).

Nếu các đánh giá lặp lại được tiến hành trong các phiên/buổi đánh giá riêng biệt rõ ràng và tùy thuộc vào thông lệ tiêu chuẩn của nhà nghiên cứu/phòng thử nghiệm, có thể sử dụng mô hình ANOVA hai chiều (tức là các hiệu ứng của thành viên hội đồng và mẫu) hoặc mô hình ANOVA ba chiều (tức là các hiệu ứng của thành viên hội đồng, phiên đánh giá và mẫu) (xem Bảng 6 và Bảng 7).

Ví dụ về ứng dụng thực tế được nêu trong Phụ lục A.

Bảng 6 – ANOVA hai chiều cho một bộ dữ liệu hoàn chỉnh (có lặp lại) cho một thuộc tính

Nguồn biến thiên	Bậc tự do	Tổng bình phương	Bình phương trung bình	Tỷ lệ F
Giữa các mẫu	$v_6 = n_p - 1$	S_6	$MS_6 = S_6/v_6$	$F = MS_6/MS_9^a$
Giữa những người đánh giá	$v_7 = n_q - 1$	S_7	$MS_7 = S_7/v_7$	$F = MS_7/MS_9^a$
Tương quan	$v_8 = (n_p - 1)(n_q - 1)$	S_8	$MS_8 = S_8/v_8$	$F = MS_8/MS_9$
Sai số	$v_9 = n_p n_q (n_r - 1)$	S_9	$MS_9 = S_9/v_9$	
Tổng cộng	$v_{10} = n_p n_q n_r - 1$	S_{10}		

n_p = số lượng mẫu;
 n_q = số lượng người đánh giá;
 n_r = số lần lặp trên mỗi mẫu;
^a Các công thức được đưa ra coi hiệu ứng của người đánh giá là cố định. Khi xem xét hiệu ứng của người đánh giá ngẫu nhiên, mẫu số của tỷ lệ F về hiệu ứng giữa các mẫu trở thành MS_8 thay vì MS_9 .

**Bảng 7 – ANOVA ba chiều cho một bộ dữ liệu hoàn chỉnh (có lặp lại)
cho một thuộc tính có tính đến hiệu ứng phiên đánh giá**

Nguồn biến thiên	Bậc tự do	Tổng bình phương	Bình phương trung bình	Tỷ lệ F
Giữa các mẫu	$v_6 = n_p - 1$	S_6	$MS_6 = S_6/v_6$	$F = MS_6/MS_{12}^a$
Giữa những người đánh giá	$v_7 = n_q - 1$	S_7	$MS_7 = S_7/v_7$	$F = MS_7/MS_{12}^a$
Giữa các phiên đánh giá	$v_{11} = n_s - 1$	S_{11}	$MS_{11} = S_{11}/v_{11}$	$F = MS_{11}/MS_{12}^a$
Tương quan	$v_8 = (n_p - 1)(n_q - 1)$	S_8	$MS_8 = S_8/v_8$	$F = MS_8/MS_{12}$
Sai số	$v_{12} = n_p n_q (n_r - 1) - (n_s - 1)$	S_{12}	$MS_{12} = S_{12}/v_{12}$	
Tổng cộng	$v_{10} = n_p n_q n_r - 1$	S_{10}		

n_p = số lượng mẫu;
 n_q = số lượng người đánh giá;
 n_r = số lần lặp trên mỗi mẫu;
 n_s = số phiên đánh giá;
^a Các công thức được đưa ra coi hiệu ứng của người đánh giá là cố định. Khi xem xét hiệu ứng của người đánh giá ngẫu nhiên, mẫu số của tỷ lệ F cho hiệu ứng giữa các mẫu trở thành MS_8 thay vì MS_{12} .

6.4 Năng lực của toàn hội đồng – Diễn giải kết quả thống kê

6.4.1 Phân biệt thuộc tính trọng yếu

Cần xác định tỷ lệ các thuộc tính trọng yếu đã được phân biệt ở mức có nghĩa như dự kiến. Đối với mỗi thuộc tính, điều này được biểu thị bằng sự dao động có nghĩa giữa các mẫu ở mức alpha 0,05 trong bảng ANOVA đối với một bộ dữ liệu hoàn chỉnh (xem Bảng 6 và Bảng 7). Tỷ lệ các thuộc tính trọng yếu

được phân biệt ở mức có nghĩa càng cao thì năng lực của hội đồng cảm quan càng cao. Hội đồng cần được huấn luyện thêm về các thuộc tính trọng yếu chưa được phân biệt ở mức có nghĩa như dự kiến.

6.4.2 Sự đồng thuận ở cấp độ hội đồng

Hội đồng không đồng thuận khi có bất kỳ người đánh giá nào không đồng thuận với những người còn lại trong hội đồng (xem 6.5.4).

Hội đồng không đồng thuận nếu sự tương tác giữa mẫu và người đánh giá trong ANOVA có ý nghĩa ở mức alpha 0,05.

Mức độ đồng thuận của hội đồng có quan hệ nghịch với sự tương tác, s_i , như được nêu trong Công thức (1):

$$s_i = \sqrt{\frac{MS_8 - MS_9}{n_r}} \quad \text{hoặc} \quad s_i = \sqrt{\frac{MS_8 - MS_{12}}{n_r}} \quad (1)$$

Xem Bảng 6 và Bảng 7.

Số lượng thuộc tính trọng yếu có sự tương tác có ý nghĩa giữa mẫu và người đánh giá được xác định dựa trên ANOVA (xem Bảng 6 và Bảng 7) cho mỗi thuộc tính. Càng nhiều thuộc tính trọng yếu mang lại sự tương tác có ý nghĩa thì hội đồng cảm quan hoạt động càng kém ổn định. Nếu sự tương tác là có ý nghĩa thì bản chất của sự tương tác cần được nghiên cứu ở cấp độ thành viên hội đồng và có hành động cần thiết. Ví dụ: nếu một thành viên hội đồng không đồng thuận về sự khác biệt giữa các sản phẩm được những người còn lại trong hội đồng đánh giá thì thành viên hội đồng này cần được huấn luyện lại.

Bản chất của sự tương tác thường được nghiên cứu bằng cách dựng đồ thị điểm trung bình sản phẩm của người đánh giá. Một lựa chọn khác là sử dụng mô hình người đánh giá hỗn hợp (MAM) [6][10].

6.4.3 Độ lặp lại của hội đồng

Độ lặp lại của hội đồng có thể được ước tính từ độ lặp lại của từng thành viên hội đồng. Độ lặp lại này quan hệ nghịch với sai số, s_e , nêu trong Công thức (2):

$$s_e = \sqrt{MS_9} \quad \text{hoặc} \quad s_e = \sqrt{MS_{12}} \quad (2)$$

tùy thuộc vào mô hình đã chọn (có hoặc không có hiệu ứng phiên đánh giá).

Xem Bảng 6 và Bảng 7.

$$S_R = \sqrt{s_e^2 + s_a^2 + s_{sess}^2 + s_{a \times sess}^2 + s_{prod \times sess}^2}$$

6.5 Năng lực của từng thành viên hội đồng – Diễn giải kết quả thống kê

6.5.1 Khả năng phân biệt của người đánh giá

Khả năng phân biệt được đo bằng tỷ lệ các thuộc tính quan trọng dự kiến được phân biệt ở mức có nghĩa. Đối với mỗi thuộc tính, điều này được biểu thị bằng sự biến thiên có ý nghĩa “giữa các mẫu” ở mức alpha 0,05 trong bảng ANOVA (xem Bảng 4 và Bảng 5). Tỷ lệ các thuộc tính trọng yếu được phân biệt ở mức có nghĩa càng cao thì năng lực của người đánh giá càng cao. Người đánh giá cần được huấn luyện thêm về các thuộc tính trọng yếu dự kiến mà chưa được phân biệt ở mức có nghĩa.

6.5.2 Độ lặp lại của người đánh giá

Độ lặp lại của người đánh giá quan hệ nghịch với sai số của họ, s_e , như nêu trong Công thức (3):

$$s_e = \sqrt{MS_2} \quad \text{hoặc} \quad s_e = \sqrt{MS_5} \quad (3)$$

tùy thuộc vào mô hình đã chọn (có hoặc không có hiệu ứng phiên đánh giá).

Xem Bảng 4 và Bảng 5.

6.5.3 Sự nhất quán của người đánh giá

Sự nhất quán của người đánh giá quan hệ nghịch với SD của độ chệch được tính từ mỗi mẫu.

(Đối với người đánh giá j , độ chệch đối với mẫu i là chênh lệch giữa giá trị trung bình của người đánh giá đối với mẫu và trung bình của hội đồng đối với mẫu, $\bar{Y}_{ji} - \bar{Y}_{i..}$. Xem Bảng 3.)

Khi năng lực của người đánh giá thiếu sự nhất quán, biểu đồ phân tán điểm số của người đánh giá so với điểm trung bình của hội đồng, cùng với phân tích hồi quy và tương quan, cho biết liệu sự không nhất quán là ngẫu nhiên hay có một mẫu cho thấy cách sử dụng thang đo khác với những người còn lại trong hội đồng.

6.5.4 Sự đồng thuận giữa những người đánh giá

Một hội đồng không đồng nhất khi có ít nhất một người đánh giá không đồng thuận với những người còn lại trong hội đồng.

Điều này có thể được phát hiện bởi:

- người đánh giá có độ chệch đáng kể;
- sai lệch của người đánh giá lớn hơn đáng kể so với toàn hội đồng;
- hệ số tương quan giữa điểm của người đánh giá và điểm trung bình của hội đồng là rất nhỏ hoặc là

số âm;

- độ dốc đường hồi quy của điểm của người đánh giá trên điểm trung bình của hội đồng là khác biệt có ý nghĩa so với 1 hoặc hệ số chặn khác biệt có ý nghĩa về mặt thống kê so với 0, hoặc cả hai.

Sự đồng thuận giữa những người đánh giá quan hệ nghịch với tương tác giữa những người đánh giá, s_a , như nêu trong Công thức (4):

$$s_a = \sqrt{\frac{MS_8 - MS_9}{n_q n_r}} \quad (4)$$

Cần kiểm tra mức độ không đồng thuận giữa những người đánh giá bằng cách sử dụng tỷ lệ F "giữa những người đánh giá" và so sánh nó với các giá trị F được tra bảng theo bậc tự do liên quan. Nếu sự khác nhau giữa hai tỷ lệ là có nghĩa thì có bằng chứng rõ ràng cho thấy có vấn đề về sự nhất quán của hội đồng và cần được giải quyết. Nếu sự khác nhau giữa hai tỷ lệ là không có nghĩa thì cũng không đảm bảo rằng không có vấn đề gì, bởi vì mức độ không đồng thuận giữa những người đánh giá có thể bị che khuất bởi độ lặp lại thấp (sai số s_e cao hơn dự kiến).

6.5.5 Độ chệch – Cách sử dụng thang đo khác nhau

Độ chệch đáng kể theo ANOVA của người đánh giá có thể chỉ ra rằng những người đánh giá sử dụng thang đo theo những cách khác nhau.

Trong hầu hết các trường hợp, không biết được giá trị "thực" nào và độ chệch tổng thể đối với người đánh giá được coi là chênh lệch giữa giá trị trung bình của người đánh giá đó và giá trị trung bình của hội đồng.

Độ chệch của người đánh giá j được nêu trong Công thức (5):

$$\bar{Y}_{.j} - \bar{Y}_{..} \quad (5)$$

Người đánh giá có thể sử dụng thang đo [xem TCVN 5090 (ISO 4121)] theo nhiều cách khác nhau. Trong việc sử dụng thang đo "phổ quát", cường độ của từng thuộc tính được đánh giá liên quan đến kiến thức của người đánh giá về sự thay đổi cảm quan tổng thể có thể trải nghiệm đối với một loại sản phẩm cụ thể. Các hội đồng làm việc trên một hoặc chỉ một vài loại sản phẩm thường phát triển phương pháp này hơn. Trong việc sử dụng thang đo "tương đối", hệ quy chiếu được người đánh giá sử dụng để đánh giá cường độ có liên quan đến sự thay đổi cảm quan được thể hiện bởi bộ sản phẩm trong một phép thử nhất định. Cách tiếp cận này có nhiều khả năng được sử dụng bởi các hội đồng làm việc trên nhiều loại sản phẩm. Để giúp giảm độ chệch do thang đo, điều quan trọng là phải đảm bảo rằng cách tiếp cận theo thang đo được nhất quán trong một hội đồng.

6.6 Các vấn đề về năng lực

6.6.1 Yêu cầu chung

Các vấn đề về năng lực sau khi được xác định có thể được liệt kê và các buổi huấn luyện được lên kế hoạch phù hợp.

6.6.2 Hội đồng cảm quan

Các buổi huấn luyện có thể được tổ chức cho toàn hội đồng về những thuộc tính gây ra vấn đề.

6.6.3 Thành viên hội đồng

Đối với các vấn đề cụ thể liên quan đến năng lực của từng thành viên hội đồng, có thể thảo luận riêng về các vấn đề đó trên cơ sở từng người theo cách trung lập hoặc tích cực, ví dụ: cung cấp dữ liệu ở cấp độ cá nhân so với mức trung bình của hội đồng làm phản hồi. Sau đó có thể thực hiện các buổi huấn luyện cho toàn hội đồng.

6.7 Thiết kế thử nghiệm để theo dõi năng lực theo thời gian

Nếu một nghiên cứu được hoạch định để đánh giá sự nhất quán của một hội đồng theo thời gian thì một phiên mỗi tháng trong khoảng thời gian một năm sẽ cung cấp đủ dữ liệu. Mỗi phiên đánh giá cần sử dụng thiết kế cân bằng nêu trong 6.2.3. Việc xem xét năng lực theo thời gian có thể được sử dụng để xác định mô hình, ví dụ: sự lệch hướng của hội đồng hoặc cải thiện năng lực sau các buổi huấn luyện lại.

7 Quy trình giám sát liên tục thông qua lập profile sản phẩm thông thường

7.1 Lựa chọn thuộc tính

Quy trình này giống như đối với việc đo lường năng lực thông qua quy trình chuyên dụng (xem Điều 6). Tuy nhiên, do không có sự khác biệt được xác định trước nên các thuộc tính được hội đồng phân biệt ở mức có nghĩa đối với một profile nhất định thì nên được sử dụng làm thước đo chính để kiểm tra năng lực của từng thành viên hội đồng. Các thuộc tính ghi lại không có sự khác biệt đáng kể thì không được sử dụng một cách đáng tin cậy để kiểm tra tính nhất quán, vì việc thiếu sự đồng thuận trong hội đồng và giữa những người đánh giá có thể là do các sản phẩm rất giống nhau về những đặc điểm đó.

7.2 Phân tích thống kê

Các phân tích thống kê tương tự như các phân tích được mô tả cho việc đo lường năng lực thông qua một quy trình chuyên dụng (xem Điều 6), ngoại trừ hiệu ứng của người đánh giá phải luôn được coi là ngẫu nhiên.

7.3 Theo dõi năng lực theo thời gian

Nếu dữ liệu từ một số phiên đánh giá định kỳ đã có sẵn thì chúng có thể được phân tích để cho thấy mọi thay đổi xảy ra theo thời gian. Việc xem xét năng lực theo thời gian có thể được sử dụng để xác định mô hình, ví dụ: cải thiện năng lực sau các buổi huấn luyện lại hoặc sự lệch hướng của hội đồng.

7.4 Phân tích thống kê dữ liệu theo thời gian

Việc phân tích dữ liệu tổng thể qua nhiều phiên đánh giá cần được thực hiện bằng các thước đo ANOVA lặp lại. Trên thực tế, những người đánh giá giống nhau có thể không có mặt ở tất cả các phiên và cần sử dụng tùy chọn mô hình tuyến tính chung của ANOVA để có được các ước tính không bị chệch về độ chệch của từng người đánh giá cũng như các thông số khác và thành phần phương sai.

Đối với hội đồng cam quan, có thể thu được ước tính a) và b).

- Sự nhất quán của hội đồng có thể được ước tính từ phiên đánh giá (xem Bảng 7), nếu dữ liệu trên các mẫu đối chứng giống hệt nhau đã được thu thập qua chuỗi phiên đánh giá.
- Sự nhất quán giữa những người đánh giá: khi có độ chệch giữa những người đánh giá, sự tương tác giữa người đánh giá và phiên đánh giá sẽ đo lường mức độ ổn định của chúng.

Đối với mỗi người đánh giá, có thể thu được ba ước tính đối với từng thuộc tính:

- Độ chệch tổng thể: mức trung bình, qua các lần lặp lại và/hoặc qua các phiên đánh giá, về chênh lệch giữa điểm của người đánh giá và điểm trung bình tương ứng của toàn hội đồng.
- Sự nhất quán: quan hệ nghịch với sự thay đổi của độ chệch giữa các phiên đánh giá.
- Độ lặp lại: biến thiên giữa điểm số của các mẫu giống hệt nhau, được xác định bằng cách gộp các ước tính SD của sai lệch (phần dư) từ mỗi phiên đánh giá.

$$s_R = \sqrt{s_{res}^2 + s_{a \times p}^2 + s_p^2}$$

7.5 Phân tích thống kê profile hoàn chỉnh

Các phương pháp phân tích thống kê được mô tả ở trên được áp dụng riêng cho từng thuộc tính để đánh giá năng lực của hội đồng và người đánh giá đối với từng thuộc tính/câu hỏi mà họ phải trả lời.

Ngoài ra, để có được cách nhìn tổng quan về toàn bộ dữ liệu, có thể sử dụng các kỹ thuật đa chiều như phân tích thành phần chính (PCA), phân tích phân biệt (DA) và phân tích Procrustes tổng quát (GPA). Những cách tiếp cận này có thể được sử dụng trong quá trình xác nhận năng lực hoặc giám sát liên tục. Để biết thêm chi tiết về các phương pháp phân tích đa biến này, xem Tài liệu tham khảo [5], [9] và [11].

Phụ lục A

(tham khảo)

Ví dụ về ứng dụng thực tế

A.1 Lập bảng dữ liệu

Tại một phiên, bốn người đánh giá đã cho điểm trên một thuộc tính với ba lần lặp lại của sáu mẫu. Bảng A.1 nêu kết quả của ví dụ này.

CHÚ THÍCH: Đây là một ví dụ minh họa. Thông thường sẽ có nhiều hơn bốn người đánh giá tham gia.

Bảng A.1 – Bảng dữ liệu về kết quả của người đánh giá

Mẫu	Người đánh giá								Trung bình
	Người đánh giá 1		Người đánh giá 2		Người đánh giá 3		Người đánh giá 4		
	Điểm số	Trung bình	Điểm số	Trung bình	Điểm số	Trung bình	Điểm số	Trung bình	
1	8		5		6		9		7,50
	8	8,3	8	7,3	7	6,0	8	8,3	
	9		9		5		8		
2	6		6		5		7		6,17
	8	7,0	7	5,7	4	5,3	7	6,7	
	7		4		7		6		
3	4		5		4		5		4,25
	5	4,7	2	3,3	3	4,0	5	5,0	
	5		3		5		5		
4	6		6		4		6		4,92
	6	5,7	4	5,3	2	3,3	5	5,3	
	5		6		4		5		
5	4		3		4		4		3,92
	5	4,0	2	3,0	4	4,3	5	4,3	
	3		4		5		4		
6	5		4		5		7		5,33
	6	5,7	2	4,3	4	5,0	5	6,3	
	6		7		6		7		
Trung bình	5,89		4,83		4,67		6,00		5,35

A.2 Phân tích thống kê

Các bảng ANOVA được trình bày chi tiết cho ví dụ này trong Bảng A.2, Bảng A.3, Bảng A.4 và Bảng A.5.

Bảng A.2 – ANOVA cho bộ dữ liệu hoàn chỉnh (người đánh giá = hiệu ứng cố định)

Nguồn biến thiên	Bậc tự do	Tổng bình phương	Bình phương trung bình	Tỷ lệ F
Giữa các mẫu	5	104,90	20,98	16,39 ^a
Giữa những người đánh giá	3	26,04	8,68	6,79 ^a
Tương quan	15	16,04	1,07	0,84 ^b
Sai lệch (phần dư)	48	61,33	1,28	
Tổng cộng	71	208,31		
^a Có nghĩa ở mức $\alpha = 0,05$;				
^b Không có nghĩa ở mức $\alpha = 0,05$.				

Bảng A.3 – Phân tích phương sai – Thành viên hội đồng

Nguồn biến động	Bậc tự do	Người đánh giá							
		Người đánh giá 1		Người đánh giá 2		Người đánh giá 3		Người đánh giá 4	
		MS	F	MS	F	MS	F	MS	F
Giữa các mẫu	5	7,42	13,36 ^a	7,83	2,66 ^b	2,80	2,40 ^b	6,13	13,80 ^a
Sai lệch	12	0,56		2,94		1,17		0,44	
SD của sai lệch, s		0,75		1,71		1,08		0,67	
^a Có nghĩa ở mức $\alpha = 0,05$;									
^b Không có nghĩa ở mức $\alpha = 0,05$.									

Bảng A.4 – Độ chệch riêng lẻ và SD của sai lệch

Người đánh giá	Độ chệch	SD của sai lệch
1	5,89 – 5,35 = +0,54	0,75
2	4,83 – 5,35 = –0,52	1,71
3	4,67 – 5,35 = –0,68	1,08
4	6,00 – 5,35 = +0,65	0,67
CHÚ THÍCH: Độ chệch là sự chênh lệch giữa giá trị trung bình của người đánh giá và giá trị trung bình tổng thể, cả hai đều được nêu trong Bảng A.1.		

Bảng A.5 – Các độ lệch mẫu riêng lẻ

Mẫu	Người đánh giá			
	1	2	3	4
1	0,83	−0,17	−1,50	0,83
2	0,83	−0,50	−0,83	0,50
3	0,42	−0,92	−0,25	0,75
4	0,75	0,42	−1,58	0,42
5	0,08	−0,92	0,42	0,42
6	0,33	−1,00	−0,33	1,00
SD, s	0,31	0,56	0,78	0,24

CHÚ THÍCH: Độ lệch riêng lẻ là chênh lệch giữa giá trị trung bình của người đánh giá đối với một mẫu và giá trị trung bình của hội đồng đối với mẫu đó, cả hai đều được nêu trong Bảng A.1.

A.3 Năng lực của toàn hội đồng – Diễn giải kết quả thống kê

Từ Bảng A.2, có thể thấy hiệu ứng "giữa các mẫu" là có ý nghĩa (ở mức alpha là 0,05), cho thấy hội đồng có thể chưa nhất quán về các sản phẩm.

Cũng từ bảng này, cũng có thể thấy rằng sự tương tác không có ý nghĩa ở mức alpha 0,05, cho thấy rằng sự thiếu nhất quán giữa các thành viên hội đồng về sự khác biệt giữa các sản phẩm là không đáng kể.

Tỷ lệ F "giữa những người đánh giá" là có ý nghĩa cho thấy rằng những người đánh giá cho điểm trung bình khác nhau (trên tất cả các sản phẩm). Mức độ biến thiên trong các điểm trung bình của người đánh giá có thể được mô tả bởi SD của người đánh giá.

Trong ví dụ này, các kết luận tương tự sẽ đạt được với hiệu ứng người đánh giá ngẫu nhiên thay vì hiệu ứng người đánh giá cố định. Một số khác biệt về cách diễn giải có thể xảy ra khi sự tương tác giữa sản phẩm*người đánh giá là có ý nghĩa.

A.4 Năng lực của từng thành viên hội đồng – Diễn giải kết quả thống kê

A.4.1 Yêu cầu chung

Người đánh giá thứ 2 và thứ 3 có SD của sai lệch cao nhất (xem Bảng A.4), cho thấy độ lặp lại thấp hơn giữa các lần lặp lại của cùng một mẫu so với Người đánh giá 1 và 4.

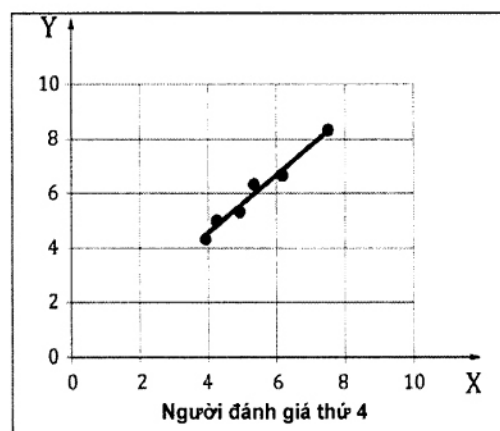
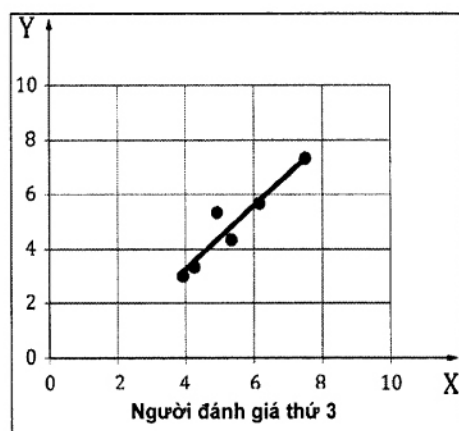
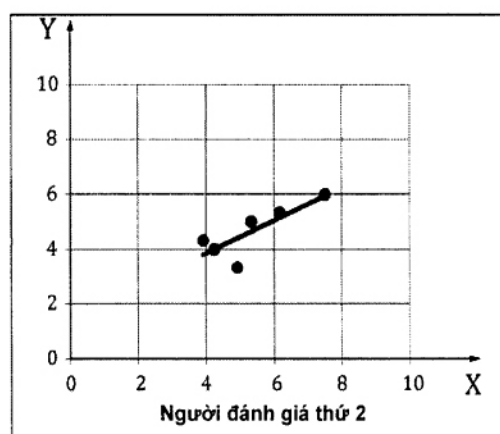
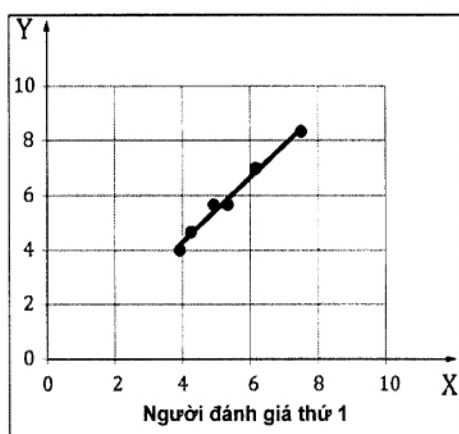
Trung bình, người đánh giá thứ 3 cũng có độ lệch âm cao, cho thấy xu hướng cho điểm thấp hơn so với những người còn lại trong hội đồng. Người đánh giá này cũng kém nhất quán so với những người

khác, thay đổi từ 1,58 dưới mức trung bình của hội đồng đến 0,42 trên mức trung bình của hội đồng, với giá trị SD lớn nhất (0,78).

Người đánh giá thứ 4 có độ chệch dương cao +0,65 nhưng nhất quán vì SD của độ chệch chỉ là 0,24. Vì người đánh giá thứ nhất và thứ 4 đều đồng thuận cao và có độ biến thiên thấp, nên có khả năng điểm số của họ đáng tin cậy và điểm trung bình của hội đồng đã bị người đánh giá thứ 2 và thứ 3 hạ xuống, do đó "độ chệch" của người đánh giá thứ 4 không phải là nguyên nhân gây lo ngại.

A.4.2 Thống kê hồi quy và tương quan

Hình A.1 cho biết điểm của mỗi người đánh giá về sáu sản phẩm được dựng đồ thị theo điểm trung bình của hội đồng.



Chú dẫn

X điểm trung bình của hội đồng

Y điểm của người đánh giá

Hình A.1– Điểm của mỗi người đánh giá so với điểm trung bình của hội đồng

TCVN 14253:2024

Trong ví dụ này, không có điểm "đúng". Giá trị trung bình của hội đồng được sử dụng làm điểm tham chiếu cho mỗi người đánh giá.

Đồ thị lý tưởng là đồ thị thể hiện sự đồng thuận hoàn toàn giữa người đánh giá và giá trị trung bình của hội đồng, với các điểm gần với độ dốc $b = 1,00$ và điểm giao $a = 0,00$. Hệ số tương quan phải ở mức gần $+1,00$.

Thông kê hồi quy và tương quan cho bốn người đánh giá được nêu trong Bảng A.6.

Bảng A.6 – Thông kê hồi quy và tương quan

Tham số	Người đánh giá			
	1	2	3	4
Tương quan	0,99	0,95	0,81	0,99
Độ dốc, b	1,18	1,16	0,59	1,07
Hệ số chặn x , a	-0,42	-1,36	1,49	0,29

Người đánh giá 4 dường như là người tốt nhất, với hệ số tương quan gần bằng 1, độ dốc gần bằng 1 và hệ số chặn nhỏ nhất.

Người đánh giá 3 có độ dốc nhỏ, cho thấy việc sử dụng thang đo hẹp hơn so với những người đánh giá khác.

Người đánh giá 2 có hệ số chặn âm, biểu thị độ chệch âm.

A.5 Các vấn đề bổ sung về năng lực

A.5.1 Yêu cầu chung

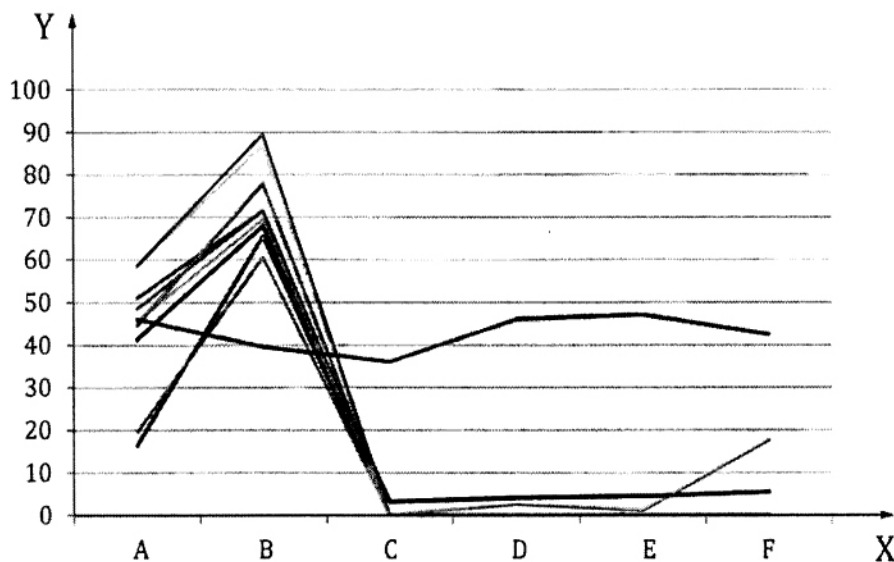
Biểu đồ đường thẳng có thể hữu ích để phát hiện các vấn đề cần phân tích thêm.

A.5.2 Thành viên hội đồng

Ba ví dụ để so sánh năng lực của từng thành viên trong một hội đồng được thể hiện trên Hình A.2 đến Hình A.4.

Hình A.2 thể hiện tình huống trong đó nhìn chung có sự đồng thuận cao về việc tách mẫu đối với tất cả người đánh giá trừ một người. Người đánh giá 10 có ít sự phân biệt giữa các mẫu. Những người đánh giá còn lại thể hiện sự đồng thuận cao đối với tất cả các mẫu ngoại trừ Mẫu A.

Thuộc tính 1

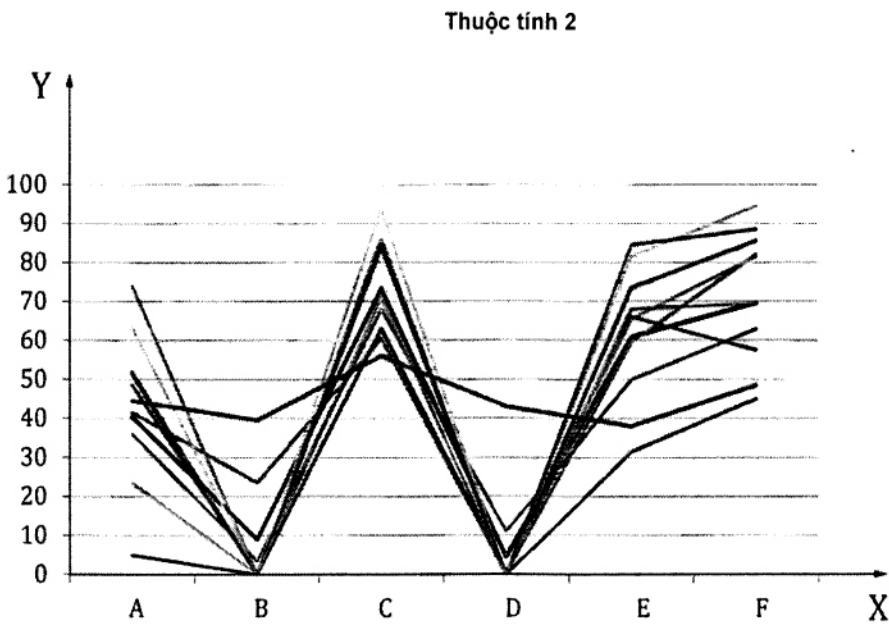


Chú dẫn

X	mẫu	_____	5
Y	điểm trung bình	_____	6
_____	điểm trung bình của hội đồng	_____	7
_____	1	_____	8
_____	2	_____	9
_____	3	_____	10
_____	4		

Hình A.2 – Điểm của hội đồng gồm mười người đối với sáu mẫu trên một thuộc tính (Thuộc tính 1)

Hình A.3 nêu tình huống trong đó hầu hết người đánh giá đều đồng thuận về thứ tự bố trí của các mẫu, nhưng Người đánh giá 10 có sự phân biệt kém và sử dụng thang đo trong phạm vi hẹp.



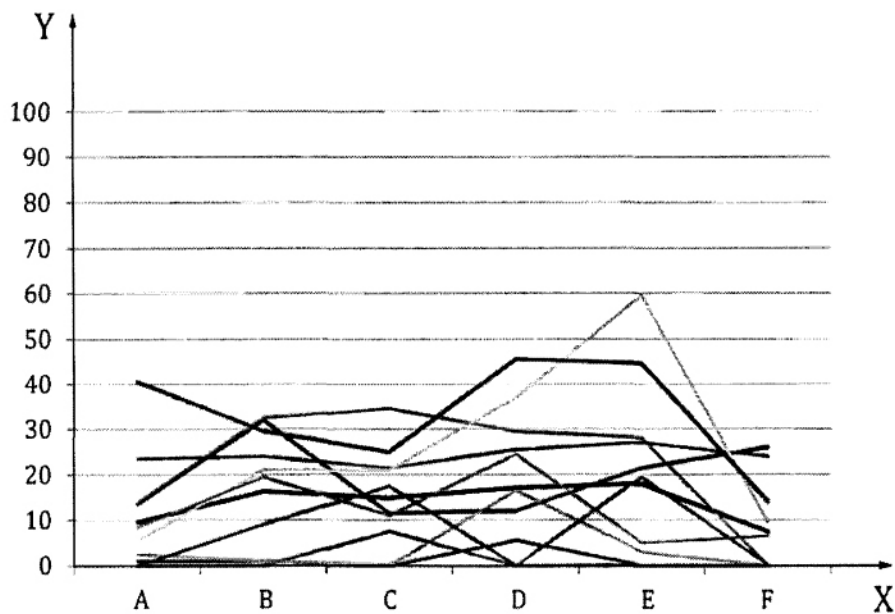
Chú dẫn

X	mẫu	5
Y	điểm trung bình	6
	điểm trung bình của hội đồng	7
	1	8
	2	9
	3	10
	4	

Hình A.3 – Điểm của hội đồng gồm mười người đối với sáu mẫu trên một thuộc tính (Thuộc tính 2)

Hình A.4 nêu tình huống trong đó tất cả người đánh giá đều có năng lực kém trong cả việc tách mẫu và sử dụng thang đo. Những người đánh giá thể hiện sự không đồng thuận ngay cả trong việc xếp loại các mẫu và hai trong số những người đánh giá cho điểm rất thấp đối với tất cả các mẫu.

Thuộc tính 3



Chú dẫn

X	mẫu	5
Y	điểm trung bình	6
	điểm trung bình của hội đồng	7
	1	8
	2	9
	3	10
	4	

Hình A.4 – Điểm của hội đồng gồm mười người đối với sáu mẫu trên một thuộc tính (Thuộc tính 3)

Thư mục tài liệu tham khảo

- [1] TCVN 5090 (ISO 4121), *Phân tích cảm quan – Hướng dẫn sử dụng các thang đo định lượng đặc trưng*
 - [2] TCVN 12389 (ISO 8586), *Phân tích cảm quan – Hướng dẫn chung để lựa chọn, huấn luyện, giám sát người đánh giá được lựa chọn và chuyên gia đánh giá cảm quan*
 - [3] TCVN 12390 (ISO 8589), *Phân tích cảm quan – Hướng dẫn chung về thiết kế phòng thử*
 - [4] TCVN 12749 (ISO 13299), *Phân tích cảm quan – Phương pháp luận – Hướng dẫn chung để thiết lập profile cảm quan*
 - [5] ARNOLD G.M., WILLIAMS A.A. The use of generalised Procrustes techniques in sensory analysis. In: Piggott J.R. (ed.) *Statistical procedures in food research*. Elsevier Applied Science, London, 1986, pp. 233–254
 - [6] BROCKHOFF P.B. SCHLICH P. SKOVGAARD I. Taking individual scaling differences into account by analyzing profile data with the Mixed Assessor Model. *Food Quality and Preference*. 2015, 39, pp. 156–166
 - [7] LEA P., NÆS T., RØDBOTTEN M. *Analysis of variance for sensory data*. Chichester: Wiley, 1997
 - [8] LUNDAHL D.S., MCDANIEL M.R. The panellist effect – fixed or random? *Journal of Sensory Studies*. 1988, 3, pp. 113–121
 - [9] NÆS T., BROCKHOFF P.B., TOMIC O. *Statistics for Sensory and Consumer Science*. John Wiley and Sons, UK, 2010
 - [10] PELTIER C. VISALLI M. SCHLICH P. Multiplicative decomposition of the scaling effect in the Mixed Assessor Model into a descriptor-specific and an overall coefficient. *Food Quality and Preference*. 2015, 48, pp. 268–273
 - [11] VARELA P., ARES G. *Novel Techniques in Sensory Characterization and Consumer Profiling*. CRC Press, 2014
 - [12] WILLIAMS E.J. Experimental designs balanced for the estimation of residual effects of treatments. *Aust. J. Sci. Res. Ser. A*. 1949, 2 pp. 149–168
-