

TCVN

TIÊU CHUẨN QUỐC GIA

TCVN 4831 : 2009

ISO 5495 : 2005

Xuất bản lần 2

**PHÂN TÍCH CẢM QUAN –
PHƯƠNG PHÁP LUẬN – PHÉP THỬ SO SÁNH CẶP ĐÔI**

Sensory analysis – Methodology – Paired comparison test

HÀ NỘI – 2009

Lời nói đầu

TCVN 4831 : 2009 thay thế TCVN 4831 : 1989;

TCVN 4831 : 2009 hoàn toàn tương đương với ISO 5495 : 2005,
Đính chính kỹ thuật 1 : 2006;

TCVN 4831 : 2009 do Ban kỹ thuật tiêu chuẩn quốc gia
TCVN/TC/F13 *Phương pháp phân tích và lấy mẫu* biên soạn,
Tổng cục Tiêu chuẩn Đo lường Chất lượng đề nghị, Bộ Khoa
học và Công nghệ công bố.

Phân tích cảm quan – Phương pháp luận – Phép thử so sánh cặp đôi

Sensory analysis – Methodology – Paired comparison test

1 Phạm vi áp dụng

Tiêu chuẩn này mô tả qui trình xác định xem có hoặc không có sự khác nhau hoặc giống nhau giữa hai mẫu về cường độ các thuộc tính cảm quan. Phép thử này đôi khi được coi như phép thử sự khác nhau có định hướng hoặc phép thử 2-AFC (Buộc phải lựa chọn thay thế). Trong thực tế, phép thử so sánh cặp đôi là phép thử lựa chọn bắt buộc giữa hai sự lựa chọn.

CHÚ THÍCH Phép thử so sánh cặp đôi là phép thử phân loại hiện hành đơn giản nhất vì chỉ liên quan đến hai mẫu.

Phương pháp này có thể áp dụng khi có sự khác nhau trong một hoặc một vài thuộc tính cảm quan riêng lẻ, có nghĩa là phương pháp này có thể xác định có hay không có sự khác nhau có thể nhận biết được liên quan đến thuộc tính đã cho và yêu cầu về hướng của sự khác nhau, nhưng phương pháp này không đưa ra bất kỳ một chỉ dẫn nào về chừng mực khác nhau này. Khi không có sự khác nhau về thuộc tính đang nghiên cứu không có nghĩa là không có sự khác nhau nào giữa hai sản phẩm.

Phương pháp này chỉ có thể áp dụng nếu các sản phẩm tương đối đồng nhất.

Phương pháp này có hiệu lực khi:

a) để xác định

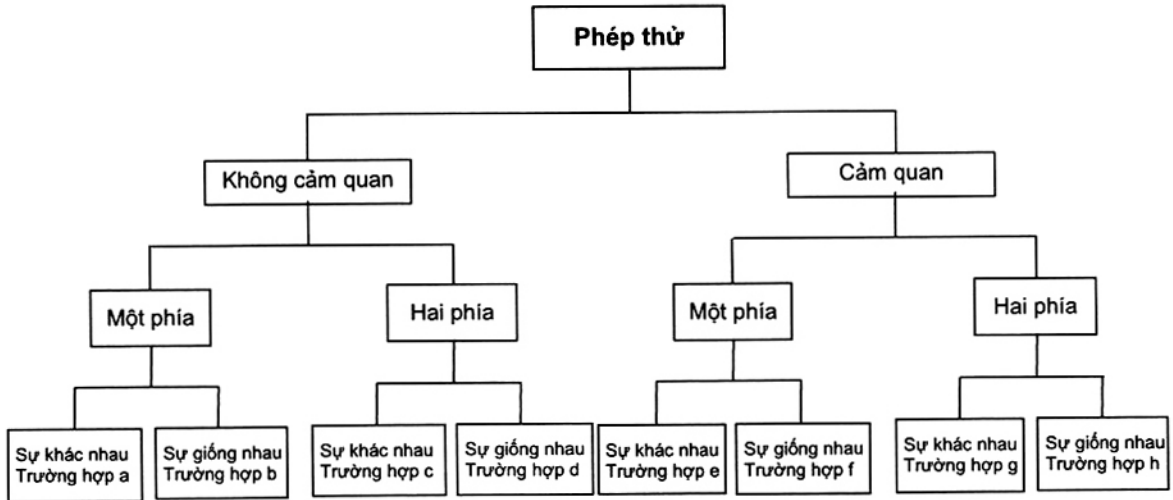
- khi có sự khác nhau có thể chấp nhận được (phép thử cặp đôi về sự khác biệt) hoặc;
- khi không có sự khác nhau có thể chấp nhận được (phép thử cặp đôi về sự giống nhau), ví dụ: có những sửa đổi về thành phần, qui trình chế biến, bao gói, vận chuyển hoặc bảo quản hoặc

b) để lựa chọn, đào tạo và giám sát người đánh giá

Trước khi bắt đầu phép thử, cần phải biết rằng đây là phép thử một phía (người giám sát phép thử biết trước hướng của sự khác nhau và giả thuyết thay thế ứng với sự tồn tại của sự khác nhau trong hướng mong đợi) hay là phép thử hai phía (người giám sát phép thử không biết trước hướng của sự

khác nhau và giả thuyết thay thế ứng với sự tồn tại của sự khác nhau trong hướng này hoặc hướng khác).

Phép thử cặp đôi có thể được sử dụng để so sánh hai sản phẩm về sự ưa thích. Các trường hợp khác nhau về cách sử dụng phép thử cặp đôi được nêu trong Hình 1.



CHÚ THÍCH Chỉ có phép thử không cảm quan là được đề cập trong tiêu chuẩn này.

Hình 1 – Các trường hợp khác nhau sử dụng phép thử so sánh cặp đôi

VÍ DỤ 1: (Trường hợp a) Việc sản xuất bánh bích qui đã được điều chỉnh để bánh được giòn hơn. Cần kiểm tra xem việc tăng độ giòn như thế có thể nhận biết được không. Do đó, cần phải kiểm tra sự khác nhau nổi bật để biết sản phẩm mới được cảm nhận là giòn hơn so với sản phẩm thông thường (sản phẩm đối chứng).

VÍ DỤ 2: (Trường hợp b) Nhà sản xuất biết rằng sản phẩm có thể chứa một lượng nhỏ của thành phần làm mất mùi sản phẩm. Do đó nhà sản xuất muốn xác định lượng tối đa có thể chấp nhận được sao cho sự khác nhau về mùi so với sản phẩm đối chứng không chứa thành phần này đủ để cảm nhận được và do đó không có sự đáng tiếc nào xảy ra tiếp theo.

VÍ DỤ 3: (Trường hợp c) Mong muốn để sản xuất món xúp mới và để so sánh hai thành phần cho vị mặn. Vì các lý do giá thành, thì thành phần có cùng nồng độ mà sẽ cho vị mặn nhất là rất cần thiết. Do đó cần cố gắng làm rõ sự khác nhau. Chưa biết trước được thành phần nào sẽ tạo được vị mặn mạnh nhất.

VÍ DỤ 4: (Trường hợp d) Nhà sản xuất chất dẻo, cụ thể là sản xuất chần bùn ô tô, để tiết kiệm, họ thay chất làm trơn thông thường bằng loại mới, nhưng không muốn công thức của chất dẻo mới có thể cảm nhận được bề mặt trơn hơn hoặc kém hơn so với loại thông thường. Vấn đề xác định ở đây là với cùng nồng độ, thì chất làm trơn mới cho cùng mức độ “trơn bề mặt” như sản phẩm thông thường. Cần chỉ ra rằng cả hai chất làm trơn là giống nhau về “trơn bề mặt”, nhưng chưa biết trước chất làm trơn nào có thể cho bề mặt trơn nhất.

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).

ISO 5492 : 1992, *Sensory analysis – Vocabulary (Phân tích cảm quan – Từ vựng)*.

ISO 6658 : 1985, *Sensory analysis – Methodology – General guidance (Phân tích cảm quan – Phương pháp luận – Hướng dẫn chung)*.

ISO 8586-1 : 1993, *Sensory analysis – General guidance for the selection, training and monitoring of assessors – Part 1: Selected assessors (Phân tích cảm quan – Hướng dẫn chung đối với việc lựa chọn, đào tạo và kiểm soát người đánh giá – Phần 1: Người đánh giá được lựa chọn)*.

ISO 8586-2 : 1994, *Sensory analysis – General guidance for the selection, training and monitoring of assessors – Part 2: Experts (Phân tích cảm quan – Hướng dẫn chung đối với việc lựa chọn, đào tạo và kiểm soát người đánh giá – Phần 2: Chuyên gia)*.

ISO 8589 : 1988, *Sensory analysis – General guidance for the design of test rooms (Phân tích cảm quan – Hướng dẫn chung đối với việc thiết kế phòng thử nghiệm)*.

3 Thuật ngữ và định nghĩa

Trong tiêu chuẩn này sử dụng các thuật ngữ, định nghĩa trong ISO 5492 và các thuật ngữ, định nghĩa sau đây:

3.1

Rủi ro α (alpha) [α (alpha) risk]

Xác suất kết luận rằng có sự khác nhau có thể chấp nhận được khi không có một sự khác nhau nào.

CHÚ THÍCH Điều này còn được gọi là sai lỗi loại I, mức có nghĩa hoặc mức dương tính giả.

3.2

Rủi ro β (beta) [β (beta) risk]

Xác suất kết luận rằng không có sự khác nhau có thể chấp nhận được khi có một sự khác nhau.

CHÚ THÍCH Điều này còn được gọi là sai lỗi loại II, mức độ âm tính giả.

3.3

Sự khác nhau (difference)

Tình huống mà trong đó các mẫu có thể phân biệt được dựa vào các thuộc tính cảm quan của chúng.

CHÚ THÍCH Tỷ lệ người đánh giá đã phát hiện sự khác nhau có thể cảm nhận được giữa các sản phẩm về thuộc tính cảm quan cần xem xét được đưa ra bằng ký hiệu p_d .

3.4

Phép thử một phía [one-sided test]

Phép thử trong đó người giám sát phép thử có kiến thức trước liên quan đến hướng của sự khác nhau.

CHÚ THÍCH Giả thuyết "không" là H_0 , là các sản phẩm không có sự khác nhau; tỷ lệ câu trả lời đúng, p , bằng 1/2. Giả thuyết thay thế H_1 , $p > 1/2$.

3.5

Phép thử hai phía [two-sided test]

Phép thử trong đó người giám sát phép thử không có kiến thức trước liên quan đến hướng của sự khác nhau.

CHÚ THÍCH Giả thuyết "không" là H_0 , là các sản phẩm không có sự khác nhau; tỷ lệ câu trả lời về một trong các mẫu, p , bằng 1/2. Giả thuyết thay thế H_1 , $p \neq 1/2$.

3.6

Trả lời đúng (correct responses)

Trả lời như mong đợi (expected responses)

Trong trường hợp phép thử một phía, số người đánh giá tham gia phép thử đã chọn đúng mẫu, được người giám sát kỳ vọng có thuộc tính cảm quan mạnh nhất.

3.7

Trả lời đồng thuận (consensual responses)

Trong trường hợp phép thử hai phía, số người đánh giá chọn mẫu A lớn hơn số người đánh giá chọn mẫu B.

CHÚ THÍCH Số người này được tính như trên vì không có bất kỳ câu trả lời nào đúng.

3.8

Sản phẩm (product)

Vật liệu cần được đánh giá

3.9

Mẫu (sample)

Đơn vị sản phẩm đã chuẩn bị, được đưa ra và được đánh giá trong quá trình đánh giá.

3.10

Độ nhạy (sensitivity)

Thuật ngữ chung được áp dụng để tổng hợp các tính năng của phép thử.

CHÚ THÍCH Trong thuật ngữ thống kê, thi độ nhạy của phép thử được xác định theo các giá trị α , β và p_d .

3.11

Tính tương đồng (similarity)

Tính hướng, trong đó sự khác nhau có thể cảm nhận được giữa các mẫu là quá nhỏ đến mức mà các mẫu đó có thể sử dụng để thay thế nhau.

4 Nguyên lý

Số lượng người đánh giá được chọn theo độ nhạy mong muốn của phép thử (Xem 6.2 và Chú thích kèm theo Bảng A.4 và Bảng A.5).

Những người đánh giá nhận được một bộ hai mẫu (một cặp). Họ sẽ chỉ định mẫu nào được coi là mạnh nhất về thuộc tính cảm quan, ngay cả khi việc lựa chọn này chỉ dựa vào dự đoán.

CHÚ THÍCH Một trong các mẫu này có thể là một mẫu đối chứng.

Số lần mà một mẫu được chọn tính được và sự khác nhau có nghĩa được xác định bằng cách so sánh với bảng thống kê có tính đến các kết quả thu được đối với mẫu dự đoán (phép thử một phía) hoặc số câu trả lời nhiều nhất thu được đối với các mẫu thử (phép thử hai phía).

5 Các điều kiện chung của phép thử

5.1 Xác định rõ ràng mục tiêu của phép thử để xác định nếu phép thử là một phía hoặc thử hai phía, nếu đó là phép thử về sự giống nhau hoặc khác nhau và phép thử nào có độ nhạy thích hợp nhất.

5.2 Tiến hành phép thử dưới các điều kiện ngăn ngừa được mọi sự trao đổi thông tin giữa những người đánh giá cho đến khi kết thúc việc đánh giá, sử dụng phòng và các trang thiết bị phù hợp với ISO 8589.

5.3 Chuẩn bị các mẫu ngoài tầm mắt của những người đánh giá và theo cách giống hệt nhau cho từng người; nghĩa là cùng loại dụng cụ, cùng loại chén.

5.4 Những người đánh giá không thể đưa ra bất kỳ kết luận nào về cường độ của thuộc tính theo cách mà các mẫu thể hiện. Ví dụ: đối với phép thử xúc giác, thì mọi sự khác nhau về bề ngoài phải tránh. Che dấu tất cả các màu khác nhau nếu mục tiêu của phép thử không liên quan đến màu sắc bằng cách sử dụng bộ lọc ánh sáng và/hoặc ánh sáng dịu. Các mẫu có thể được thể hiện liên tiếp hoặc không đồng thời trong trường hợp hơi có sự khác nhau về bề ngoài.

5.5 Mã số của các mẫu hoặc các chén đựng mẫu phải theo cách thống nhất, tốt nhất là sử dụng các số có 3 chữ số được chọn ngẫu nhiên cho mỗi phép thử. Mỗi cặp gồm hai mẫu, có mã số khác nhau. Tốt nhất là nên sử dụng các mã số khác nhau cho mỗi người đánh giá trong suốt phiên đánh giá. Tuy nhiên có thể sử dụng hai mã giống nhau cho tất cả những người đánh giá trong một phép thử, với điều kiện là mỗi mã số chỉ được sử dụng một lần cho một người đánh giá trong quá trình thử đánh giá (ví dụ: nếu trong cùng một phiên đánh giá tiến hành vài cặp thử trên các sản phẩm khác nhau).

5.6 Số lượng hoặc thể tích được cung cấp phải giống nhau đối với hai mẫu thử tạo thành một cặp, giống như tất cả các mẫu khác đối với sản phẩm đã nêu. Số lượng hoặc thể tích cần được đánh giá có thể được ấn định. Nếu không, điều này cần qui định cho người đánh giá để lấy các lượng hoặc các thể tích mẫu tương tự.

5.7 Nhiệt độ của các cặp mẫu phải bằng nhiệt độ của tất cả các mẫu khác trong dãy phép thử trên sản phẩm đã nêu. Tốt nhất là để các mẫu ở nhiệt độ mà sản phẩm thường được tiêu dùng.

5.8 Những người đánh giá phải được thông báo trước rằng họ có phải tuân thủ thủ tục đặc biệt để đánh giá các sản phẩm hay không (ví dụ: có được nuốt mẫu để thử vị hoặc thực hiện động tác thể hiện về thử xúc giác hay không) hay là họ tự do làm như họ muốn. Khi được tự do làm những gì mình thích thì họ cần phải thực hiện theo cùng phương thức đối với tất cả các mẫu.

5.9 Trong suốt quá trình đánh giá, tránh đưa thông tin về nhận biết sản phẩm, hiệu quả xử lý mong đợi hoặc sự thực hiện riêng cho đến khi kết thúc các phép thử.

6 Người đánh giá

6.1 Trình độ chuyên môn

Tất cả những người đánh giá cần có trình độ chuyên môn mức ngang nhau, mức này được chọn trên cơ sở mục tiêu của phép thử (xem ISO 8586-1 và ISO 8586-2). Kinh nghiệm và sự hiểu biết sản phẩm có thể làm tăng khả năng của người đánh giá và do vậy có thể làm tăng khả năng tìm ra sự khác nhau có nghĩa. Việc giám sát khả năng của những người đánh giá qua thời gian cho thấy độ nhạy tăng.

Tất cả những người đánh giá phải hiểu biết về cơ chế của phép thử cặp đôi (phiếu ghi điểm, nhiệm vụ và qui trình đánh giá). Ngoài ra, những người đánh giá phải có thể nhận biết về thuộc tính cảm quan là cơ sở của phép thử. Thuộc tính này phải được xác định bằng chất chuẩn hoặc bằng cách thể hiện vài mẫu có các mức cường độ thuộc tính khác nhau cần thử nghiệm.

6.2 Số lượng người đánh giá

Chọn số lượng người đánh giá sao cho thu được độ nhạy theo yêu cầu đối với phép thử (xem Bảng A.4 về phép thử một phía và Bảng A.5 về phép thử hai phía). Sử dụng số lượng lớn người đánh giá làm tăng khả năng phát hiện đúng các sai khác nhỏ giữa các sản phẩm. Tuy nhiên trong thực tế, số lượng người đánh giá thường được xác định bởi các điều kiện thiết yếu (ví dụ: thời gian thực nghiệm, số lượng người đánh giá sẵn có, lượng sản phẩm). Khi tiến hành phép thử sự khác nhau, thì số lượng người đánh giá điển hình thường từ 24 người đến 30 người. Khi thực hiện phép thử sự giống nhau thì số lượng người đánh giá thường gấp đôi (khoảng 60 người). Khi thử nghiệm về tính giống nhau, các phép đánh giá không nên lặp lại bởi cùng người đánh giá. Tuy nhiên, nếu yêu cầu các phép đánh giá lặp lại để có đủ số lượng đánh giá tổng số, thì cần cố gắng để mỗi người đánh giá thực hiện cùng một số lượng đánh giá lặp lại. Ví dụ: nếu chỉ có sẵn 10 người đánh giá, thì mỗi người đánh giá sẽ tiến hành ba phép thử cặp đôi để thu được tổng số 30 lần đánh giá.

CHÚ THÍCH Phân tích ba lần đánh giá do 10 người đánh giá thực hiện theo 30 lần đánh giá độc lập là không có hiệu lực khi thử nghiệm về tính tương đồng sử dụng Bảng A.3. tuy nhiên, phép thử sự khác nhau sử dụng Bảng A.1 và A.2 là có hiệu lực ngay cả khi thực hiện các đánh giá lặp lại [5], [6]. Một số ấn bản gần đây [1], [2] về các phép thử về sự khác biệt kếp gợi ý các cách tiếp cận thay thế về phân tích các đánh giá lặp lại.

7 Cách tiến hành

7.1 Chuẩn bị giấy và bảng chấm điểm (Xem Hình B.1, B.2 và B.3) trước khi bắt đầu phép thử để sử dụng số lượng như nhau của hai chuỗi trình bày liên tục của cả hai sản phẩm A và B.

7.2 Trình bày hai mẫu của cặp so sánh liên tục hoặc đồng thời (xem 5.4). Trong trường hợp trình bày liên tục, thì bố trí hai mẫu theo cách giống nhau đối với mỗi người đánh giá (theo hàng từ trái sang phải, từ dưới lên trên v.v...). Những người đánh giá sẽ kiểm tra hai mẫu tạo thành cặp đôi theo thứ tự trong bảng chấm điểm, nhưng những người đánh giá thường được phép đưa ra các đánh giá lặp lại của mỗi mẫu nếu muốn (dĩ nhiên là nếu bản chất của sản phẩm cho phép đánh giá lặp lại).

7.3 Cần cung cấp một bảng chấm điểm cho mỗi cặp mẫu. Nếu người đánh giá thực hiện nhiều hơn một phép thử trong quá trình đánh giá thì thu lấy bảng chấm điểm đã hoàn thiện và các mẫu chưa sử dụng trước khi thực hiện cặp mẫu tiếp theo. Đánh giá viên không được quay lại với bất kỳ mẫu thử nào trước đó, cũng như không thay đổi quyết định nào liên quan đến phép thử trước đó.

7.4 Không được có bất kỳ câu hỏi nào về ưu điểm, sự chấp nhận hoặc độ khác nhau theo sự lựa chọn của mẫu có cường độ mạnh nhất. Việc lựa chọn mà người đánh giá vừa làm xong có thể sai lệch so với câu trả lời cho mọi câu hỏi bổ sung. Các câu trả lời cho những câu hỏi đó có thể thu được qua các phép thử riêng rẽ liên quan đến ưu điểm, sự chấp nhận hoặc độ khác nhau... (xem ISO 6658). Phần "Nhận xét" đề nghị đưa ra các lý do lựa chọn mà có thể bao gồm trong các lưu ý của người đánh giá.

7.5 Phép thử so sánh cặp đôi là qui trình "lựa chọn bắt buộc"; những người đánh giá không được phép chọn "không có sự khác nhau". Khi người đánh giá phát hiện không có sự khác nhau giữa các mẫu cần được hướng dẫn chọn một trong các mẫu và chỉ ra rằng việc lựa chọn chỉ là phỏng đoán trong phần "Nhận xét" của bảng chấm điểm.

8 Phân tích và diễn giải kết quả

8.1 Khi thử về sự khác nhau

8.1.1 Trường hợp thử một phía

Sử dụng Bảng A.1 để phân tích các số liệu thu được trong phép thử cặp đôi. Nếu số câu trả lời đúng bằng hoặc lớn hơn số lượng nêu trong Bảng A.1 (tương ứng với số lượng người đánh giá và với mức rủi ro α đã chọn cho phép thử), thì kết luận rằng có sự khác nhau có thể nhận biết được giữa các mẫu (xem B.1).

Nếu muốn, tính khoảng tin cậy trên tỷ lệ những người có thể phân biệt được các mẫu. Phương pháp này được qui định trong B.5.

8.1.2 Trường hợp thử hai phía

Sử dụng Bảng A.2 để phân tích các số liệu thu được trong phép thử cặp đôi. Nếu số câu trả lời đồng thuận bằng hoặc lớn hơn số lượng nêu trong Bảng A.2 (tương ứng với số lượng người đánh giá và với mức rủi ro α đã chọn cho phép thử), thì kết luận rằng có sự khác nhau có thể nhận biết được giữa các mẫu (xem B.3).

Nếu muốn, tính khoảng tin cậy trên tỷ lệ những người có thể phân biệt được các mẫu. Phương pháp này được qui định trong B.5.

8.2 Khi thử về sự tương đồng ¹⁾

8.2.1 Trường hợp thử một phía

Sử dụng Bảng A.3 để phân tích các số liệu thu được trong phép thử cặp đôi. Nếu số câu trả lời đúng nhỏ hơn hoặc bằng số lượng nêu trong Bảng A.3 (tương ứng với số lượng người đánh giá, với mức rủi ro β và giá trị p_d đã chọn cho phép thử), thì kết luận rằng có sự khác nhau không đáng kể giữa các mẫu (xem B.2). Nếu các kết quả của phép thử này được so sánh với kết quả của một phép thử khác thì cần chọn cùng một giá trị p_d cho tất cả các phép thử.

Nếu muốn, tính khoảng tin cậy trên tỷ lệ những người có thể phân biệt được các mẫu. Phương pháp này được qui định trong B.5.

Không có kết luận nào đưa ra với các số tối đa các câu trả lời đúng dưới $n/2$.

8.2.2 Trường hợp thử hai phía

Sử dụng Bảng A.3 để phân tích các số liệu thu được trong phép thử cặp đôi. Nếu số câu trả lời đúng nhỏ hơn hoặc bằng số lượng nêu trong Bảng A.3 (tương ứng với số lượng người đánh giá, với mức rủi ro β và với giá trị p_d đã chọn cho phép thử), thì kết luận rằng có sự khác nhau không đáng kể giữa các mẫu (xem B.4). Nếu các kết quả của phép thử này được so sánh với kết quả của một phép thử khác thì cần chọn cùng một giá trị p_d cho tất cả các phép thử.

Nếu muốn, tính khoảng tin cậy trên tỷ lệ những người có thể phân biệt được các mẫu. Phương pháp này được qui định trong B.5.

¹⁾ Trong tiêu chuẩn, "tương đồng" (similar) không có nghĩa là "đồng nhất" (identical). Thuật ngữ này biểu thị rằng hai sản phẩm hoàn toàn giống nhau có thể sử dụng thay thế nhau. Điều này không có nghĩa là hai sản phẩm đồng nhất. Tuy nhiên, điều đó cũng có thể cho thấy giữa hai sản phẩm có sự khác nhau quá nhỏ đến mức không đáng kể trong thực tế.

9 Báo cáo thử nghiệm

Báo cáo mục đích thử nghiệm, các kết quả thu được và các kết luận. Khuyến cáo bổ sung các thông tin sau đây:

- mục đích của phép thử và bản chất của việc xử lý cần nghiên cứu;
- nhận biết đầy đủ về mẫu thử: nguồn gốc, phương pháp chuẩn bị, lượng, hình dạng, bảo quản trước khi thử, lượng mẫu cung cấp, nhiệt độ (thông tin liên quan đến mẫu cần chỉ rõ đầy đủ về bảo quản, vận chuyển và các thao tác chuẩn bị cần phải thực hiện sao cho tạo được các mẫu mà khác nhau chỉ do có sự thay đổi về điều quan tâm, nếu có);
- số lượng đánh giá viên, số câu trả lời đúng hoặc đồng thuận và kết quả đánh giá thống kê (bao gồm các giá trị α , β và p_d dùng trong phép thử);
- những người đánh giá: kinh nghiệm (trong thử đánh giá cảm quan với sản phẩm, với mẫu thử), tuổi và giới tính (xem ISO 8586-1 và ISO 8586-2);
- mọi thông tin cụ thể và các khuyến cáo đưa ra cho người đánh giá liên quan đến phép thử, cụ thể là khi xác định chính xác và các mẫu chuẩn minh họa thuộc tính cần thử và/hoặc thủ tục thử nghiệm đã được chỉ cho người đánh giá;
- môi trường thử nghiệm: thiết bị sử dụng, thực hiện đồng thời hay kế tiếp, việc nhận biết các mẫu đã được phơi ra sau khi thử nghiệm và nếu có thì theo cách nào;
- nơi, ngày thử nghiệm và tên của chủ tịch hội đồng.

10 Độ chụm và độ chệch

Vì các kết quả của các phép thử phân biệt về cảm quan phụ thuộc vào độ cảm nhận của từng cá thể, nên không thể đưa ra tuyên bố chung liên quan đến độ tái lập của các kết quả để có thể áp dụng cho tất cả những người đánh giá. Độ chụm liên quan đến số lượng cụ thể của người đánh giá tăng theo cỡ nhóm, cũng như với sự đào tạo và tiếp xúc với sản phẩm.

Vì qui trình "lựa chọn bắt buộc" được sử dụng, nên các kết quả thu được bằng phương pháp này không bị chệch với điều kiện là tuân thủ đầy đủ Điều 7.

Phụ lục A

(Qui định)

Các bảng

A.1 Xác định sự khác nhau có thể cảm nhận được hoặc tính tương đồng

Xem Bảng A.1 đến A.3.

Bảng A.1 – Số lượng tối thiểu các câu trả lời đúng cần thiết để kết luận rằng có sự khác nhau cảm nhận được, dựa trên phép thử so sánh cặp đôi một phía ^{2) 3)}

n	α					n	α				
	0,20	0,10	0,05	0,01	0,001		0,20	0,10	0,05	0,01	0,001
10	7	8	9	10	10	36	22	23	24	26	28
11	8	9	9	10	11	37	22	23	24	27	29
12	8	9	10	11	12	38	23	24	25	27	29
13	9	10	10	12	13	39	23	24	26	28	30
14	10	10	11	12	13	40	24	25	26	28	31
15	10	11	12	13	14						
						44	26	27	28	31	33
16	11	12	12	14	15	48	28	29	31	33	36
17	11	12	13	14	16	52	30	32	33	35	38
18	12	13	13	15	16	56	32	34	35	38	40
19	12	13	14	15	17	60	34	36	37	40	43
20	13	14	15	16	18						
						64	36	38	40	42	45
21	13	14	15	17	18	68	38	40	42	45	48
22	14	15	16	17	19	72	41	42	44	47	50
23	15	16	16	18	20	76	43	45	46	49	52
24	15	16	17	19	20	80	45	47	48	51	55
25	16	17	18	19	21						
						84	47	49	51	54	57
26	16	17	18	20	22	88	49	51	53	56	59
27	17	18	19	20	22	92	51	53	55	58	62
28	17	18	19	21	23	96	53	55	57	60	64
29	18	19	20	22	24	100	55	57	59	63	66
30	18	20	20	22	24						
						104	57	60	61	65	69
31	19	20	21	23	25	108	59	62	64	67	71
32	19	21	22	24	26	112	61	64	66	69	73
33	20	21	22	24	26	116	64	66	68	71	76
34	20	22	23	25	27	120	66	68	70	74	78
35	21	22	23	25	27						

CHÚ THÍCH 1 Các giá trị trong bảng là chính xác bởi vì chúng được dựa trên sự phân bố nhị thức. Đối với các giá trị n không nằm trong Bảng này thì có thể thu được giá trị gần đúng theo cách sau đây: Số lượng tối thiểu câu trả lời đúng (x) bằng số nguyên gần nhất lớn hơn $x = (n + 1)/2 + z\sqrt{0,25n}$ trong đó z thay đổi theo hàm số mức có nghĩa như sau: 0,84 đối với $\alpha = 0,20$; 1,28 đối với $\alpha = 0,10$; 1,64 đối với $\alpha = 0,05$; 2,33 đối với $\alpha = 0,01$; 3,09 đối với $\alpha = 0,001$.

CHÚ THÍCH 2 Các giá trị $n < 18$ thường không được khuyến cáo để sử dụng trong các phép thử cặp đôi khác nhau.

²⁾ Các giá trị đưa ra trong Bảng này tính được từ công thức đúng của phân bố nhị thức đối với thông số $p = 0,5$ với n lần lặp lại theo phần mềm SAS trong [4].

³⁾ Các giá trị tương ứng với số câu trả lời đúng tối thiểu yêu cầu để có ý nghĩa tại mức α đã nêu (cột dọc) đối với số lượng người đánh giá tương ứng, n (hàng ngang). Loại bỏ việc khẳng định "không có sự khác nhau" nếu số lượng câu trả lời đúng lớn hơn hoặc bằng giá trị nêu trong bảng.

Bảng A.2 – Số lượng tối thiểu các câu trả lời đồng thuận cần thiết để kết luận rằng có sự khác nhau cảm nhận được, dựa trên phép thử so sánh cặp đôi hai phía ^{2), 3)}

n	α					n	α				
	0,20	0,10	0,05	0,01	0,001		0,20	0,10	0,05	0,01	0,001
10	8	9	9	10		36	23	24	25	27	29
11	9	9	10	11	11	37	23	24	25	27	29
12	9	10	10	11	12	38	24	25	26	28	30
13	10	10	11	12	13	39	24	26	27	28	31
14	10	11	12	13	14	40	25	26	27	29	31
15	11	12	12	13	14						
						44	27	28	29	31	34
16	12	12	13	14	15	48	29	31	32	34	36
17	12	13	13	15	16	52	32	33	34	36	39
18	13	13	14	15	17	56	34	35	36	39	41
19	13	14	15	16	17	60	36	37	39	41	44
20	14	15	15	17	18						
						64	38	40	41	43	46
21	14	15	16	17	18	68	40	42	43	46	48
22	15	16	17	18	19	72	42	44	45	48	51
23	16	16	17	19	20	76	45	46	48	50	53
24	16	17	18	19	21	80	47	48	50	52	56
25	17	18	18	20	21						
						84	49	51	52	55	58
26	17	18	19	20	22	88	51	53	54	57	60
27	18	19	20	21	23	92	53	55	56	59	63
28	18	19	20	22	23	96	55	57	59	62	65
29	19	20	21	22	24	100	57	59	61	64	67
30	20	20	21	23	25						
						104	60	61	63	66	70
31	20	21	22	24	25	108	62	64	65	68	72
32	21	22	23	24	26	112	64	66	67	71	74
33	21	22	23	25	27	116	66	68	70	73	77
34	22	23	24	25	27	120	68	70	72	75	79
35	22	23	24	26	28						

CHÚ THÍCH 1 Các giá trị trong bảng là chính xác bởi vì chúng được dựa trên sự phân bố nhị thức. Đối với các giá trị n không nằm trong Bảng này thì có thể thu được giá trị gần đúng theo cách sau đây: Số lượng tối thiểu câu trả lời đúng (x) bằng số nguyên gần nhất lớn hơn $x = (n + 1)/2 + z\sqrt{0,25n}$ trong đó z thay đổi theo hàm số mức có nghĩa như sau: 1,28 đối với $\alpha = 0,20$; 1,64 đối với $\alpha = 0,10$; 1,96 đối với $\alpha = 0,05$; 2,58 đối với $\alpha = 0,01$; 3,29 đối với $\alpha = 0,001$.

CHÚ THÍCH 2 Các giá trị $n < 18$ thường không được khuyến cáo để sử dụng trong các phép thử cặp đôi khác nhau.

Bảng A.3 – Số lượng tối đa các câu trả lời đúng hoặc câu trả lời đồng thuận cần thiết để kết luận rằng hai mẫu giống nhau, dựa trên phép thử so sánh cặp đôi^{4) 5)}

n	β	p_a					n	β	p_a				
		10 %	20 %	30 %	40 %	50 %			10 %	20 %	30 %	40 %	50 %
18	0,001	-	-	-	-	-	48	0,001	-	-	-	-	25
	0,01	-	-	-	-	-		0,01	-	-	-	25	28
	0,05	-	-	-	-	9		0,05	-	-	25	27	30
	0,10	-	-	-	9	10		0,10	-	-	26	28	31
	0,20	-	-	9	10	11		0,20	-	25	27	30	33
24	0,001	-	-	-	-	-	54	0,001	-	-	-	-	29
	0,01	-	-	-	-	12		0,01	-	-	-	29	32
	0,05	-	-	-	12	13		0,05	-	-	28	31	34
	0,10	-	-	12	13	14		0,10	-	27	30	32	35
	0,20	-	-	13	14	15		0,20	-	28	31	34	37
30	0,001	-	-	-	-	-	60	0,001	-	-	-	-	33
	0,01	-	-	-	-	16		0,01	-	-	-	33	36
	0,05	-	-	-	16	17		0,05	-	-	32	35	38
	0,10	-	-	15	17	18		0,10	-	30	33	36	40
	0,20	-	15	16	18	20		0,20	-	32	35	38	41
36	0,001	-	-	-	-	-	66	0,001	-	-	-	-	37
	0,01	-	-	-	18	20		0,01	-	-	33	36	40
	0,05	-	-	18	20	22		0,05	-	-	35	39	43
	0,10	-	-	19	21	23		0,10	-	34	37	40	44
	0,20	-	18	20	22	24		0,20	-	35	39	42	46
42	0,001	-	-	-	-	21	72	0,001	-	-	-	37	40
	0,01	-	-	-	21	24		0,01	-	-	36	40	44
	0,05	-	-	21	23	26		0,05	-	-	39	43	47
	0,10	-	-	22	25	27		0,10	-	37	41	44	48
	0,20	-	22	24	26	28		0,20	-	39	42	46	50

⁴⁾ Các giá trị đưa ra trong Bảng này nhờ chương trình tính toán các khoảng tin cậy từ công thức đúng của phân bố nhị thức trong [7].

⁵⁾ Các giá trị này ứng với số lượng tối đa câu trả lời đúng hoặc câu trả lời đồng thuận yêu cầu cho "tính tương đồng" ở các mức p_a , β và n đã chọn. Chấp nhận "không có sự khác nhau" ở mức $100(1 - \beta)$ % mức tin cậy nếu số câu trả lời đúng hoặc câu trả lời đồng thuận ít hơn hoặc bằng giá trị nêu trong bảng.

Bảng A.3 (kết thúc)

n	β	P_d					n	β	P_d				
		10 %	20 %	30 %	40 %	50 %			10 %	20 %	30 %	40 %	50 %
78	0,001	-	-	-	40	44	108	0,001	-	-	54	59	65
	0,01	-	-	40	44	48		0,01	-	-	57	63	69
	0,05	-	39	43	47	51		0,05	-	55	61	67	72
	0,10	-	40	44	48	53		0,10	-	57	63	68	74
	0,20	-	42	46	50	54		0,20	54	60	65	71	76
84	0,001	-	-	-	44	48	114	0,001	-	-	57	63	69
	0,01	-	-	43	48	53		0,01	-	-	61	67	73
	0,05	-	42	46	51	55		0,05	-	59	65	71	77
	0,10	-	44	48	52	57		0,10	-	61	67	72	79
	0,20	-	46	50	54	59		0,20	57	63	69	75	81
90	0,001	-	-	-	48	53	120	0,001	-	-	61	67	73
	0,01	-	-	47	52	57		0,01	-	-	65	71	78
	0,05	-	45	50	55	60		0,05	-	62	68	75	81
	0,10	-	47	52	56	61		0,10	-	64	70	77	83
	0,20	45	49	54	58	63		0,20	60	67	73	79	85
96	0,001	-	-	-	52	57	126	0,001	-	-	64	70	77
	0,01	-	-	50	56	61		0,01	-	-	68	75	82
	0,05	-	49	54	59	64		0,05	-	66	72	79	85
	0,10	-	50	55	60	66		0,10	-	68	74	81	87
	0,20	48	53	58	62	68		0,20	64	70	76	83	89
102	0,001	-	-	-	55	61	132	0,001	-	-	67	74	81
	0,01	-	-	54	59	65		0,01	-	65	72	79	86
	0,05	-	52	57	63	68		0,05	-	69	76	83	90
	0,10	-	54	59	64	70		0,10	-	71	78	85	92
	0,20	51	56	61	67	72		0,20	67	73	80	87	94

CHÚ THÍCH 1 Các giá trị trong bảng là chính xác bởi vì chúng được dựa trên sự phân bố nhị thức. Đối với các giá trị n không nằm trong Bảng này, tính giới hạn tin cậy trên $100(1 - \beta) \%$ đối với p_d , như sau:

$$[2(x/n) - 1] + 2 \times z_\beta \times \sqrt{(nx - x^2)/n^3}$$

trong đó x là số câu trả lời đúng hoặc đồng thuận, n là số người đánh giá và z_β thay đổi như sau: 0,84 đối với $\beta = 0,20$; 1,28 đối với $\beta = 0,10$; 1,64 đối với $\beta = 0,05$; 2,33 đối với $\beta = 0,01$; 3,09 đối với $\beta = 0,001$. Nếu giá trị tính được thấp hơn giới hạn đã chọn trước của p_d thì công bố rằng các mẫu là tương đồng ở mức ý nghĩa β .

CHÚ THÍCH 2 Các giá trị $n < 30$ thường không được khuyến cáo để sử dụng trong các phép thử cặp đôi tương đồng.

CHÚ THÍCH 3 Các giá trị tương ứng với số câu trả lời đúng dưới $n/2$ không được đề cập trong Bảng này. Chúng được mã hoá bằng kí hiệu -.

A.2 Cách tiếp cận thống kê để xác định số lượng người đánh giá trên cơ sở Bảng A.4 (phép thử một phía) và A.5 (phép thử hai phía)

Độ nhạy thống kê của phép thử phụ thuộc vào ba giá trị: rủi ro α , rủi ro β và tỷ lệ tối đa cho phép của "phân biệt" p_d ⁶⁾. Trước khi bắt đầu phép thử, chọn các giá trị α , β và p_d sử dụng các hướng dẫn sau đây:

Theo nguyên tắc chung, kết quả có ý nghĩa về thống kê đối với rủi ro α :

- giữa 10 % và 5 % (0,10 đến 0,05) cho thấy có bằng chứng (mức độ nhẹ) về sự khác nhau;
- giữa 5 % và 1 % (0,05 đến 0,01) cho thấy có bằng chứng (mức độ vừa) về sự khác nhau;
- giữa 1 % và 0,1 % (0,01 đến 0,001) cho thấy có bằng chứng (mức độ cao) về sự khác nhau và
- dưới 0,1 % (<0,001) cho thấy có bằng chứng chắc chắn về sự khác nhau.

Đối với rủi ro β , thì mức độ của bằng chứng rằng không có sự khác nhau được đánh giá bằng cách sử dụng cùng chuẩn cứ như qui định ở trên (thay "có" bằng "không").

Tỷ lệ tối đa cho phép của "phân biệt" p_d được chia thành ba dải:

- $p_d < 25\%$ thể hiện các giá trị nhỏ;
- $25\% < p_d < 35\%$ thể hiện các giá trị trung bình và
- $p_d > 35\%$ thể hiện các giá trị lớn.

Chọn số lượng người đánh giá sao để thu được mức độ nhạy yêu cầu cho phép thử. Nhận biết trong Bảng A.4 phản tương ứng với giá trị p_d được chọn và cột tương ứng với giá trị β được chọn. Số lượng tối đa yêu cầu đối với người đánh giá được để trong dòng tương ứng với giá trị α được chọn. Cách khác, có thể sử dụng Bảng A.4 để xây dựng một tập hợp các giá trị α , β và p_d mà có thể cho độ nhạy chấp nhận được trong khi số lượng người đánh giá còn lại nằm trong các giới hạn thực tế. Cách tiếp cận này được nêu chi tiết trong [4].

⁶⁾ Trong tiêu chuẩn này, xác suất của một câu trả lời đúng, p_c , được tính theo công thức $p_c = 1 \times p_d + (1/2) \times (1 - p_d)$, trong đó p_d là tỉ lệ người đánh giá có thể phân biệt hai mẫu. Mô hình lượng hoá tâm lí của quá trình ra quyết định của người đánh giá, ví dụ như mô hình Thurstone-Ura [3] có thể được áp dụng cho phép thử cặp đôi.

Bảng A.4 – Số lượng người đánh giá cần thiết cho phép thử cặp đôi một phía^{7) 8)}

α		β					
		0,50	0,20	0,10	0,05	0,01	0,001
0,50	$p_d = 50\%$	– ^a	–	–	9	22	33
0,20		–	12	19	26	39	58
0,10		–	19	26	33	48	70
0,05		13	23	33	42	58	82
0,01		35	40	50	59	80	107
0,001		38	61	71	83	107	140
0,50		$p_d = 40\%$	–	–	9	20	33
0,20	–		19	30	39	60	94
0,10	14		28	39	53	79	113
0,05	18		37	53	67	93	132
0,01	35		64	80	96	130	174
0,001	61		95	117	135	176	228
0,50	$p_d = 30\%$		–	–	23	33	59
0,20		–	32	49	68	110	166
0,10		21	53	72	96	145	208
0,05		30	69	93	119	173	243
0,01		64	112	143	174	235	319
0,001		107	172	210	246	318	412
0,50		$p_d = 20\%$	–	23	45	67	133
0,20	21		77	112	158	253	384
0,10	46		115	168	214	322	471
0,05	71		158	213	268	392	554
0,01	141		252	325	391	535	726
0,001	241		386	479	556	731	944
0,50	$p_d = 10\%$		–	75	167	271	539
0,20		81	294	451	618	1006	1555
0,10		170	461	658	861	1310	1905
0,05		281	620	866	1092	1583	2237
0,01		550	1007	1301	1582	2170	2927
0,001		961	1551	1908	2248	2937	3812

^a Các ô trống ứng với các trường hợp không cho bất kỳ một sự quan tâm nào (các giá trị cao đối với α và β có tính đến giá trị p_d được chọn).

⁷⁾ Các giá trị đưa ra trong Bảng này được lấy từ [4] hoặc tính được từ công thức đúng của qui luật phân bố nhị thức đối với thông số $p = 0,5$ với n câu trả lời dựa vào phần mềm SAS trong tài liệu tham khảo này.

⁸⁾ Các giá trị ứng với số lượng tối thiểu của người đánh giá cần thiết để thực hiện phép thử cặp đôi với độ nhạy được xác định bằng các giá trị p_d , α và β . Nhận biết trong bảng phần ứng với giá trị đã chọn p_d , cột dọc ứng với giá trị β được chọn. Đọc số lượng tối thiểu người đánh giá trên hàng ngang ứng với giá trị α được chọn.

Bảng A.5 – Số lượng người đánh giá cần thiết cho phép thử cặp đôi hai phía ^{8) 9)}

α		β					
		0,50	0,20	0,10	0,05	0,01	0,001
0,50	$p_d = 50\%$	– ^a	–	–	23	33	52
0,20		–	19	26	33	48	70
0,10		–	23	33	42	58	82
0,05		17	30	42	49	67	92
0,01		26	44	57	66	87	117
0,001		42	66	78	90	117	149
0,50	$p_d = 40\%$	–	–	25	33	54	86
0,20		–	28	39	53	79	113
0,10		18	37	53	67	93	132
0,05		25	49	65	79	110	149
0,01		44	73	92	108	144	191
0,001		48	102	126	147	188	240
0,50	$p_d = 30\%$	–	29	44	63	98	156
0,20		21	53	72	96	145	208
0,10		30	69	93	119	173	243
0,05		44	90	114	145	199	276
0,01		73	131	164	195	261	345
0,001		121	188	229	267	342	440
0,50	$p_d = 20\%$	–	63	98	135	230	352
0,20		46	115	168	214	322	471
0,10		71	158	213	268	392	554
0,05		101	199	263	327	455	635
0,01		171	291	373	446	596	796
0,001		276	425	520	604	781	1010
0,50	$p_d = 10\%$	–	240	393	543	910	1423
0,20		170	461	658	861	1310	1905
0,10		281	620	866	1092	1583	2237
0,05		390	801	1055	1302	1833	2544
0,01		670	1167	1493	1782	2408	3203
0,001		1090	1707	2094	2440	3152	4063

^a Các ô trống ứng với các trường hợp không cho bất kỳ một sự quan tâm nào (các giá trị cao đối với α và β có tính đến giá trị p_d được chọn).

⁹⁾ Các giá trị đưa ra trong Bảng này tính được từ công thức đúng của qui luật phân bố nhị thức đối với thông số $p = 0,5$ với n câu trả lời dựa vào phạm mềm SAS trong [4].

Phụ lục B

(Tham khảo)

Các ví dụ

B.1 Ví dụ 1 – Phép thử so sánh cặp đôi một phía để khẳng định rằng có sự khác nhau liên quan đến cường độ của thuộc tính giữa hai mẫu

B.1.1 Nội dung

Theo một số ghi nhận của khách hàng, một số sửa đổi về công nghệ đã được thực hiện để tạo ra bánh qui giòn hơn so với sản phẩm thường dùng. Trước khi tiến hành thử nghiệm trên diện rộng với sự tham gia của khách hàng, thì bộ phận phát triển phải chắc chắn rằng các thay đổi công nghệ đưa ra có được hiệu quả như mong muốn. Với mong muốn để hạn chế rủi ro về sự khác nhau. Mặt khác, vì có khả năng thay đổi công nghệ nên sẵn sàng chấp nhận rủi ro cao về sự không phát hiện được sự khác nhau hiện có.

B.1.2 Mục tiêu phép thử

Mục tiêu là để khẳng định rằng sản phẩm mới là giòn hơn thực sự. Do đó đây là trường hợp đối với phép thử một phía.

B.1.3 Số lượng người đánh giá

Để tránh cho bộ phận phát triển sản phẩm không đưa ra kết luận sai về vị, thì người giám sát phân tích cảm quan đề nghị một ngưỡng α bằng 0,05, phần trăm người đánh giá phát hiện có sự khác nhau p_d bằng 30 % và β bằng 0,50. Do đó cần tham khảo Bảng A.4 và tìm thấy là cần ít nhất 30 người đánh giá.

B.1.4 Kiểm soát phép thử

Ba mươi đĩa bánh qui "A" (kiểm chứng) và 30 đĩa bánh qui "B" (nguyên mẫu) được mã hoá bằng các số thống nhất ngẫu nhiên. Đối với 15 người đánh giá, các sản phẩm được thể hiện theo thứ tự AB, đối với 15 người đánh giá còn lại sản phẩm được thể hiện theo BA. Bảng chấm điểm mẫu được nêu trong Hình B.1.

B.1.5 Phân tích và diễn giải các kết quả

Hai mốt người đánh giá chỉ định mẫu B là giòn hơn. Đối chiếu với Bảng A.1 trong hàng ngang tương ứng với $n = 30$ và cột dọc $\alpha = 0,05$, có thể thấy rằng 20 câu trả lời theo hướng mong đợi đáp ứng được sự công bố rằng 2 mẫu khác nhau đáng kể.

B.1.6 Báo cáo và kết luận

Người phân tích cảm quan ghi lại trong báo cáo mẫu ban đầu giòn hơn, với hội đồng ($n = 30, x = 21$) ở mức ý nghĩa 5 %. Do đó, bánh qui có thể được sản xuất với qui trình mới để đáp ứng sự mong muốn của khách hàng.

Phép thử cặp đôi	
Tên:.....	Mã số đánh giá..... Ngày, tháng.....
Hướng dẫn:	
Thử ném hai mẫu bắt đầu từ mẫu phía bên tay trái của bạn. Chỉ rõ mã số của mẫu cho thấy giòn nhất vào phần trống dưới đây. Nếu bạn không chắc chắn, bạn có thể chỉ rõ trong phần "Nhận xét" rằng đó chỉ là phỏng đoán.	
Mẫu giòn nhất là:	
Nhận xét nếu có:.....	
.....	
.....	
.....	

Hình B.1 – Bảng chấm điểm cho ví dụ 1

B.2 Ví dụ 2 – Phép thử so sánh cặp đôi một phía để khẳng định hai mẫu giống nhau liên quan đến thuộc tính đã cho

B.2.1 Nội dung

Nhà sản xuất biết rằng sản phẩm có thể chứa một lượng nhỏ của thành phần làm mất mùi thơm thảo mộc của sản phẩm. Do đó nhà sản xuất muốn xác định lượng tối đa có thể chấp nhận được sao cho sự khác nhau về mùi so với sản phẩm đối chứng không chứa thành phần này (T) đủ để cảm nhận được và do đó không có sự đáng tiếc nào xảy ra tiếp theo.

B.2.2 Mục tiêu phép thử

Mục tiêu là để xác định lượng tối đa của thành phần sao cho sự khác của mùi thảo mộc với sản phẩm đối chứng không chứa thành phần này đủ để cảm nhận được và do đó không có sự đáng tiếc nào xảy ra tiếp theo.

B.2.3 Số lượng người đánh giá

Nhà sản xuất hy vọng sự rõ ràng hợp lý của các yêu cầu liên quan đến lượng cho phép của thành phần làm mất mùi thơm thảo mộc. Do đó, trong phép thử này thì rủi ro không phát hiện được sự khác

nhau trong mùi thảo mộc (β) cần được giữ càng thấp càng tốt. Rủi ro α kết luận sai về mùi khi khác nhau thực sự không có, là không quan trọng lắm, vì điều này chỉ có thể dẫn đến yêu cầu bảo toàn lớn hơn. Do đó, β cố định = 0,05, $\alpha = 0,50$ và phần trăm người đánh giá phát hiện có sự khác nhau $p_d = 20\%$. Do đó, nhà sản xuất tra cứu Bảng A.4 và thấy cần đến ít nhất là 67 người đánh giá. Tuy nhiên, trên bảng tra cứu A.3, cho thấy rằng đối với các giá trị được chọn β và p_d , thì số người đánh giá tối đa cần đến là 78 người để có thể sử dụng bảng (trong trường hợp dưới 78 người, thì số lượng tối đa đúng được đưa ra hoặc các câu trả lời đồng thuận là cơ hội thấp, nghĩa là $n/2$ và do đó không nêu trong bảng). Do đó nhà sản xuất quyết định thành lập hội đồng mới gồm 78 người đánh giá.

B.2.4 Kiểm soát phép thử

Nồng độ đích C được xác định, có tính đến các phép thử ban đầu và sự hiểu biết trước đó. Hai dung dịch được chuẩn bị và mỗi dung dịch được chia thành 78 cốc được mã hoá bằng các số ngẫu nhiên thống nhất. Đối với 39 người đánh giá, các sản phẩm được trình bày theo trật tự TC, còn đối với 39 người đánh giá còn lại thì các sản phẩm được trình bày theo trật tự CT. Bảng chấm điểm mẫu được nêu trong Hình B.1 nhưng câu hỏi so sánh các mẫu được thay bằng "mùi thảo mộc".

B.2.5 Phân tích và diễn giải các kết quả

78 người đánh giá tham gia vào phép thử này. Bốn mươi một người đánh giá chỉ ra mẫu C thơm hơn. Trong Bảng A.3 có thể cho thấy rằng đối với $n = 78$, $p_d = 20$ và $\beta = 0,05$ thì giá trị tối đa kết luận rằng các mẫu giống nhau bằng 39. Vì số câu trả lời đúng thu được trong quá trình thử nghiệm cao hơn giá trị này, nên người phân tích báo cáo rằng không thể đưa ra kết luận về sự giống nhau của mùi giữa các mẫu.

B.2.6 Báo cáo và kết luận

Người phân tích cảm quan ghi lại trong báo cáo rằng nồng độ C là quá mạnh để có thể chấp nhận thành giá trị dung sai cho phép và khuyến cáo thử nghiệm nồng độ hơi thấp hơn.

B.3 Ví dụ 3 – Phép thử so sánh cặp đôi hai phía để khẳng định rằng có sự khác nhau liên quan đến cường độ của thuộc tính giữa hai mẫu

B.3.1 Nội dung

Nhà chế biến xúp muốn xác định hai thành phần cơ bản của natri cho cường độ của vị mặn. Thành phần này sẽ được chọn từ công thức hình thành sản phẩm mới vì nó có thể được sử dụng ở nồng độ nhẹ hơn và sẽ rẻ hơn (cả hai sản phẩm đòi hỏi cùng một giá cho một kilogam). Nếu không có sự khác nhau đáng kể, thì các thành phần khác có thể được thử nghiệm.

B.3.2 Mục tiêu phép thử

Mục tiêu là để xác định xem trong hai thành phần thì thành phần nào cho cường độ của vị mặn mạnh hơn khi có cùng nồng độ. Do đó, đây là trường hợp đối với phép thử hai phía.

B.3.3 Số lượng người đánh giá

Người phân tích cảm quan muốn chắc chắn 95 % rằng tỷ lệ những người đánh giá cao sẽ có thể cảm nhận được sự khác nhau, do đó α cố định = 0,05 và $p_d = 50$ %. Tuy nhiên, khi kết luận sai rằng không có sự khác nhau sẽ dẫn đến các chi phí bổ sung vì sẽ phải thử nghiệm thêm các thành phần khác. Tiếp đó, người phân tích cảm quan cố định β bằng 0,10. Trên Bảng A.5 có thể thấy rằng cần đến ít nhất 42 người đánh giá. Do đó, quyết định thành lập hội đồng gồm 44 thành viên.

B.3.4 Kiểm soát phép thử

Chuẩn bị hai dãy xúp A và xúp B, chúng khác nhau hoàn toàn về thành phần tạo vị mặn. Hai mẫu này được chuẩn bị ở trạng thái nóng đưng trong các bát gốm đã được mã hoá bằng các số ngẫu nhiên thống nhất. Đối với 22 người đánh giá, các sản phẩm được thể hiện theo thứ tự AB, đối với 22 người đánh giá còn lại sản phẩm được thể hiện theo BA. Bảng chấm điểm mẫu được nêu trong Hình B.2.

B.3.5 Phân tích và diễn giải các kết quả

Bốn mươi bốn người đánh giá tham gia vào phép thử này. Ba mươi hai người đánh giá chỉ ra mẫu A mặn hơn và 12 người đánh giá chỉ ra mẫu B mặn hơn. Trong Bảng A.2 có thể cho thấy rằng đối với $n = 44$, cột dọc $\alpha = 0,05$ thì có thể thấy rằng số cao nhất của người đánh giá thu được cho A hoặc B phải cao hơn hoặc bằng 29 để có thể công bố 2 mẫu khác nhau đáng kể. Bảng A.2 cũng cho thấy giá trị lớn hơn hoặc bằng 31 có nghĩa là sự khác nhau đáng kể ở ngưỡng 1 % có thể được kết luận.

B.3.6 Báo cáo và kết luận

Người phân tích cảm quan ghi lại trong báo cáo rằng thành phần A cảm nhận được mặn hơn so với thành phần B ở ngưỡng 1 %. Do đó, trong tương lai thành phần A được chấp nhận để sản xuất.

Phép thử cặp đôi	
Mã số phép thử: 845-2003	
Người đánh giá số 14: Tên.....	Ngày, tháng.....
Hướng dẫn:	
Bạn hãy nếm hai mẫu từ trái sang phải. Chọn mẫu mặn nhất và nhận dạng bằng cách đánh dấu chéo vào ô tương ứng.	
842 <input type="checkbox"/>	376 <input type="checkbox"/> Nhận xét:
.....	
.....	
Nếu bạn có nhận xét liên quan đến lý do bạn chọn hoặc các đặc trưng của mẫu, bạn có thể ghi vào dưới phần "Nhận xét".	

Hình B.2 – Bảng chấm điểm cho Ví dụ 3

B.4 Ví dụ 4 – Phép thử so sánh cặp đôi hai phía để khẳng định hai mẫu giống nhau liên quan đến thuộc tính đã cho

B.4.1 Nội dung

Nhà sản xuất chất dẻo đã sử dụng, cụ thể là sản xuất chần bùn ô tô, để tiết kiệm, họ thay chất làm trơn thông thường bằng loại mới, nhưng không muốn công thức của chất dẻo mới này có thể cảm nhận được bề mặt trơn hơn hoặc kém hơn so với loại thông thường.

B.4.2 Mục tiêu phép thử

Mục tiêu là để xác định xem ở cùng một nồng độ thì liệu chất làm trơn mới có cho "bề mặt trơn" như sản phẩm thông thường hay không.

B.4.3 Số lượng người đánh giá

Nhà sản xuất hy vọng rằng thành phần mới cho cùng một mức "bề mặt trơn" cảm nhận được như loại chất làm trơn thông thường. Do đó, trong phép thử này thì rủi ro không phát hiện được sự khác nhau về độ trơn bề mặt (β) cần được giữ càng thấp càng tốt. Rủi ro α kết luận sai về "bề mặt trơn" khi sự khác nhau thực sự không có, là không quan trọng lắm, vì điều này chỉ có thể dẫn đến việc giữ lại chất bôi trơn thường dùng đắt hơn. Do đó, β cố định = 0,05, $\alpha = 0,10$ và phần trăm người đánh giá phát hiện có sự khác nhau $p_d = 30\%$. Do đó, nhà sản xuất tra cứu Bảng A.5 và thấy cần đến ít nhất là 119 người đánh giá. Để cân đối các thứ tự thể hiện, quyết định thành lập hội đồng mới gồm 120 người đánh giá.

B.4.4 Kiểm soát phép thử

Chuẩn bị hai tấm chất dẻo A và B, chúng khác nhau hoàn toàn về chất làm trơn. Hai tấm mẫu này được cho vào hai hộp được mã hoá bằng các số ngẫu nhiên thống nhất. Đối với 60 người đánh giá, các sản phẩm được thể hiện theo thứ tự AB, đối với 60 người đánh giá còn lại sản phẩm được thể hiện theo BA. Bảng chấm điểm mẫu được nêu trong Hình B.3.

B.4.5 Phân tích và diễn giải các kết quả

Hội đồng gồm 120 người đánh giá tham gia vào phép thử này. Trong đó có 67 thành viên chỉ ra mẫu A có bề mặt trơn hơn và 53 thành viên chỉ ra mẫu B có bề mặt trơn hơn. Trong Bảng A.3 có thể cho thấy rằng đối với $n = 120$, hàng ngang với $\beta = 0,05$ và $p_d = 30\%$ thì tìm được giá trị 68. Vì số câu trả lời cao nhất thu được trong quá trình thử nghiệm bằng 67 thấp hơn giá trị này (68), nên có thể đưa ra kết luận rằng có sự giống nhau giữa hai mẫu ở khoảng tin cậy 95% và ở mức p_d đã chọn cho phép thử.

B.4.6 Báo cáo và kết luận

Người phân tích cảm quan ghi lại trong báo cáo rằng chất làm trơn mới cho mức độ trơn bề mặt giống như chất làm trơn thường dùng và trong tương lai chất làm trơn mới này có thể được chọn để sản xuất.

Phép thử cặp đôi	
Tên..... Mã số của người đánh giá:..... Ngày, tháng.....	
Hướng dẫn:	
Bạn hãy so sánh 2 mẫu trên quan điểm đặc trưng về độ trơn bề mặt. Tắm có bề mặt trơn là tắm không cho thấy cản trở nào khi sờ tay lên bề mặt của tắm này. Bạn hãy sờ vào cả hai sản phẩm, bắt đầu từ tắm phía bên trái và quan sát như sau: chuyển động lòng bàn tay áp nhẹ từ trái sang phải.	
Chọn mẫu có bề mặt trơn nhất và nhận dạng bằng cách đánh dấu chéo vào ô tương ứng.	
192 <input type="checkbox"/>	526 <input type="checkbox"/>

Hình B.3 – Bảng chấm điểm cho Ví dụ 4

B.5 Ví dụ 5 – Khoảng tin cậy đối với các phép thử cặp đôi**B.5.1 Nội dung**

Nếu muốn, người phân tích có thể tính khoảng tin cậy trên tỷ lệ của những người đánh giá thực tế để có thể phân biệt 2 mẫu (nghĩa là được hiệu chỉnh về khả năng). Các phép tính được dựa trên khoảng gần đúng của phân bố chuẩn như sau, trong đó x là số câu trả lời theo hướng mong đợi (phép thử một phía) hoặc số lượng câu trả lời đồng thuận (phép thử hai phía) và n là tổng số câu trả lời:

– p_c (tỷ lệ câu trả lời mong đợi hoặc câu trả lời đồng thuận) = x / n

– \hat{p}_d (tỷ lệ phân biệt khác nhau) = $2 p_c - 1$

– s_d (sai số chuẩn của \hat{p}_d) = $2\sqrt{p_c(1-p_c)/n} = 2\sqrt{(n \times x - x^2)/n^3}$

– giới hạn tin cậy trên = $\hat{p}_d + z_\alpha s_d$

– giới hạn tin cậy dưới = $\hat{p}_d - z_\alpha s_d$

trong đó z_α là giá trị tới hạn của phân bố chuẩn đã được chuẩn hoá được đưa ra trong Bảng sau đây:

Mức	Phép thử	
	z_α một phía	z_α hai phía
80 %	0,84	1,28
90 %	1,28	1,64
95 %	1,64	1,96
99 %	2,33	2,58
99,9 %	3,10	3,29

B.5.2 Phân tích và diễn giải các kết quả

Trong Ví dụ 3 (phép thử sự khác nhau so sánh cặp đôi hai phía), thu được số liệu như sau: $n = 44$, $x = 32$, $\alpha = 0,05$ và $\beta = 0,10$:

- $p_c = x/n = 32/44 = 0,73$
- \hat{p}_d (tỷ lệ phân biệt khác nhau) = $2 p_c - 1 = 2 \times 0,73 - 1 = 0,45$
- s_d (sai số chuẩn của \hat{p}_d) = $2\sqrt{(n \times x - x^2)/n^3} = 2 \times \sqrt{(32 \times 44 - 32^2)/44^3} = 0,134$
- 95 % giới hạn tin cậy trên = $\hat{p}_d + z_\alpha s_d = 0,45 + 1,96 \times 0,134 = 0,71$
- 95 % giới hạn tin cậy dưới = $\hat{p}_d - z_\alpha s_d = 0,45 - 1,96 \times 0,134 = 0,19$

Do đó, người phân tích cảm quan có thể chắc chắn 95 % rằng ít nhất 19 % người dân có thể phân biệt được sự khác nhau giữa các mẫu. Kết quả này phù hợp với kết luận đưa ra trong Ví dụ 3, cho thấy mẫu A mạnh hơn vì cho thấy khoảng tin cậy không chứa giá trị "không" và nằm ngoài giới hạn ban đầu ấn định cho phép thử ($p_c = 50 \%$).

Trong Ví dụ 4 (phép thử tính tương đồng so sánh cặp đôi hai phía), thu được số liệu như sau: $n = 120$, $\alpha = 0,10$ và $\beta = 0,05$:

- $p_c = x/n = 67/120 = 0,56$
- \hat{p}_d (tỷ lệ phân biệt khác nhau) = $2 p_c - 1 = 2 \times 0,56 - 1 = 0,12$

TCVN 4831 : 2009

$$- s_d (\text{sai số chuẩn của } \hat{p}_d) = 2\sqrt{(n \times x - x^2)/n^3} = 2 \times \sqrt{(120 \times 67 - 67^2)/120^3} = 0,09$$

$$- 95 \% \text{ giới hạn tin cậy trên} = \hat{p}_d + z_{\alpha} s_d = 0,12 + 1,96 \times 0,09 = 0,29$$

$$- 95 \% \text{ giới hạn tin cậy dưới} = \hat{p}_d - z_{\alpha} s_d = 0,12 - 1,96 \times 0,09 = -0,06$$

Do đó, người phân tích cảm quan có thể chắc chắn 95 % rằng tỷ lệ có thể phân biệt được sự khác nhau giữa các mẫu không quá 29 %.

Tính đồng thời, các khoảng tin cậy cho phép sai số 5 % đối với các giới hạn trên và giới hạn dưới, do đó, người phân tích cảm quan có thể chắc chắn 90 % rằng tỷ lệ phân biệt sự khác nhau từ 0 % đến 29 %. Giá trị này thấp hơn các giới hạn ấn định ban đầu ($p_d = 30 \%$). Do đó, người phân tích có thể kết luận rằng không có sự khác nhau có thể cảm nhận được giữa các mẫu.

Thư mục tài liệu tham khảo

- [1] ENNIS, D.M. and Bi, J. The Beta-Binomial Model: Accounting for Inter-trial Variation in Replicated Difference and Preference Tests. *Journal of Sensory Studies*, 13, No. 4,1998, pp. 389-412
 - [2] BROCKHOFF, P.B. and SCHLICH, P. Handling replications in discrimination tests. *Food Quality and Preference*, 9, No. 5, 1998, pp. 303-312
 - [3] FRIJTERS, J.E.R. Three-Stimulus Procedure in Olfactory Psychophysics: An Experimental Comparison of Thurstone-Ura and Three-Alternative Forced-Choice Models of Signal Detection Theory. *Perception and Psychophysics*, 28, No. 5,1980, pp. 390-397
 - [4] SCHLICH, P. Risk Tables for Discrimination Tests. *Food Quality and Preference*,4,1993, pp.141-151
 - [5] KUNERT, J. and MEYNER, M. On the triangle test with replications. *Food Quality and Preference*, 10,1999
 - [6] KUNERT, J. On repeated difference testing. *Food Quality and Preference*, 12, 2001, pp 385-391
 - [7] MACRAE, A. Confidence intervals for the triangular test. *Food Quality and Preference*, 6, 1995, pp. 61-67
 - [8] ISO 3534-1, *Statistics - Vocabulary and symbols - Part 1: Probability and general statistical terms*
-