

[連載講座]

官能検査 4

官能評価に用いられる統計手法

サントリー（株） 朝倉 康夫

1 はじめに

なぜ、統計手法が必要なのか？ それは、官能検査が人間を測定の手段とすることからきている。人間を計器として、その感覚にたよって検査・評価を行うものであるから、機器測定に比べて、どうしてもあいまいさが付きまとう。この原因は、まず個人毎の評価基準に差（個人差）があることによる。たとえば、香りに対する好みの違いとか、酸味、苦味にたいする感度の違いのため、判定結果に違いが生じることがある。また、個人内でも感覚が時々刻々変化しバラツキ（個人精度）が大きい。体調やその時の気分により感度が違ったり、判断尺度が変動したりすることがある。空腹時か満腹時かで感度が異なったり、いらいらして集中出来ない時、識別力が低下したりすることがよくある。また、いくつも試料を評価しているうちに感覚順応（疲労）を起こしてしまい、最初の方の試料と後の方の試料で判定基準が異なることがある。さらに先入観によって判定が甘くなったり、辛くなったりすることもある。一流ブランドのワインであることを知ったうえでテイastingすると、評価が甘くなったり、保管状態が悪かったと聞くと、評価がいつもより厳しくなったりすることがある。

このような問題を軽減し、信頼性の高いデータを得るには、まず検査員の訓練、検査環境（温度、湿度、照明、騒音など）の整備、さらに提示試料のコントロール（グラス、入味量、品温、ブライ

ンド提示など）の要件を十分に整えたうえで、統計的にデータを処理する官能検査手法の適用が必要となる。

今回は、官能検査の主な手法について、その使用目的に応じて分類し、各手法の概略を説明した。そして、もっとも基本となる2点識別法を例にとり、統計的検定の考え方を説明した。最後に、企業活動において、官能検査が適用される分野と問題の形式を示し、それぞれに対応する手法を紹介した。適切な手法選定の目安になれば幸いである。

2 官能検査手法の分類

官能検査手法の選択にあたっては、その目的に合わせて適切なものを選び出さなければ、満足な結果が得られないばかりか、時には間違った結論を引き出すこともある。まずなによりも大切なことは、何のために官能検査をするのか、官能検査によって何を知らうとしているのかという目的意識を明確にすることである。例えば、試料間の質的な差の有無を知りたいのか、差の有無だけでなく差の大きさまで知りたいのか、試料の質ではなくて、その試料に対する被験者の好みの程度を知りたいのか、また、試料の数が2個か3個以上か、などによって手法が規制される。

2-1 分析型と嗜好型

まず、官能検査は目的によって、次のような2つのタイプに大別される。第一は、人間の感覚器

表1 分析型と嗜好型官能検査の特徴

	分析型	嗜好型
目的	品質差の検出 品質の尺度化・定量化 品質特性描写 など	嗜好の分析
検査員	専門家 人数は少数 識別能力は大 客観的判断	一般人 人数は多数 識別能力は小 主観的判断
適用分野	品質検査 工程検査 品質改良 など	嗜好調査 イメージ調査 など

官を測定器として、検査対象となる試料の特性を測定したり、試料間の差を検出したりする場合、すなわち、試料の品質を分析機器のように客観的に捉えたい場合で、これを分析型官能検査という。分析型は、客観的判断を必要とするので、検査員は十分訓練を積んだ専門家によりおこなわれる。第二は、ワインの風味に対する好みなど、対象物に対する被験者の嗜好を調べる場合、すなわち、試料に対する各被験者の主観的品質を調べたい場合で、これを嗜好型官能検査という。嗜好型は、一般消費者（特に、その商品のターゲットとなる消費者）の嗜好を代表するようなパネルが選ばれる。表1に分析型と嗜好型の特徴をまとめた。官能検査手法を用いる場合、その目的が分析型であるのか、嗜好型であるのかによって、手法が異なること、特に有意差検定のやり方が異なることがあるので注意しなければならない。

ること、特に有意差検定のやり方が異なることがあるので注意しなければならない。

2-2 目的別分類

次に、試料間に差があるかどうか（識別）、差の大きさはどのくらいか、品質の差はどのくらいか（尺度化・定量化）、品質の特性を描写する（特性描写）といった目的にそって、よく使われる手法を分類した。表2によく使われる手法の目的別分類を示した。

3 手法の説明

3-1 識別型手法

試料間に差があるのかどうか知りたい場合に適用される。標準品と比べて品質差があるのか、試料間に嗜好差があるのかといった場合に用いられる。また、

表2 官能検査手法の分類

	分析型	嗜好型
識別	2点識別法	2点嗜好法
	3点識別法	3点嗜好法
	1:2点識別法	
尺度化・定量化	順位法	順位法
	格付け法	カテゴリー尺度法
	カテゴリー尺度法	(嗜好尺度法、 嗜好意欲尺度法)
	採点法	一対比較法
特性描写	一対比較法	
	プロファイル法 QDA法	SD法

検査員の識別能力を調べる場合にも使われる。試料間の差があるかどうか分からないくらい小さな場合に適用されることが多く、差が大きい場合には後述の尺度化・定量化の手法が適用出来る。

2点識別法・2点嗜好法

試料の種類が2種類あって、2つ1組にして提示し、質問事項（例えば、甘さの強い方はどちらか）について両者を比較させ、該当する方を選択させる方法である。

2点識別法は、客観的に差のある2つの試料間の刺激の強さを比較・判定するもので、2点嗜好法は、2点のうちどちらを好むかを判定するものである。有意性は、まぐれ当たりの確率が $1/2$ であるとして、2項分布によって検定される。2点識別法は、片側検定、2点嗜好法は、両側検定となる。

3点識別法・3点嗜好法

これも比較する試料は2種類である。そして、2つは同じ試料、1つは違う試料を3つ1組にして提示し、どれが異なる1つであるかを指摘させる方法である。3点嗜好法は、まず3点識別法で違う1個を選び出した後、次に選び出した試料と残りの試料を比較して、どちらが好きかを判定させるものである。

3点識別法の場合は、有意性はまぐれ当たりの確率が $1/3$ であるとして、2項分布による片側検定をおこなう。3点嗜好法の場合は、まず3点識別法と同じ方法で、識別の有意性を検定する。有意差があった場合のみ、正答者の回答だけを集め、嗜好差を検定する。この検定は、まぐれ当たりの確率が $1/6$ であるとして、2項分布による両側検定をおこなう。

1：2点識別法

最初に標準品Aを提示して、検査員にその特徴を記憶してもらい、次にAとBを盲試料として提示し、どちらがA（あるいはB）かを指摘させる

方法である。有意差の検定は、2点識別法とまったく同じにおこなうことが出来る。

3-2 尺度化・定量化

試料間の差の大きさの判定や、順位づけ、格づけ、評点づけをしたい場合に適用される。

順位法

数個の試料にたいして、ある特性（甘さ、好ましさなど）について、順位を付ける方法である。官能検査では、各試料に対して評点をつけるのは難しいが、順位なら付けられるといったことがよくあるが、こういう場合に適用される。

順位データの解析には、スピアマンの順位相関係数、ケンドールの順位相関係数、ケンドールの一致性係数Wとフリードマンの検定、およびクレマーの検定などの手法があるが、数点の試料間の差を調べる目的には、一致性係数Wを検定したうで、クレマーの検定をおこなう。

格づけ法

「特級、1級、2級」、「合格、不合格」などに試料を格づけする方法である。

得られたデータを分類データと見るか、順位データと見るか、採点データと見るかによって、統計的処理方法が異なる。分類データとみた場合は、 χ^2 （カイ2乗）分布による独立性の検定をおこなう。順位データとみた場合には、一致性係数Wの検定およびクレマーの検定をおこなう。採点データとみた場合には、通常の計量値と同様に分散分析法と多重比較により解析される。

採点法

「非常に強い（+3）、かなり強い（+2）、やや強い（+1）、普通（0）、やや弱い（-1）、かなり弱い（-2）、非常に弱い（-3）」とか、「良い（+2）、やや良い（+1）、どちらでもない（0）、やや悪い（-1）、悪い（-2）」などの力

テグリー尺度を使って、各試料の特性や嗜好に評点を付ける方法である。用いる尺度には、カテゴリーが全て定義されている場合、一部定義されている場合、さらに、「10点満点で評価せよ」といったまったく定義されていない場合がある。この場合、パネルは、カテゴリーの幅が一定となるように判断することが要求される。これによって、得られたデータは通常の計量値と同じように、分散分析法、多重比較などの統計処理ができる。

一対比較法

数個の試料を比較するとき、すべての試料を一度に比較評価するのが困難な場合、全体のなかから2個ずつ取り出して比較する。そしてこれらの対比較のデータをもとにして全試料を評価する方法である。データの取り方、データ処理の仕方により、シェッフエの方法、ブラッドレーの方法、サーストン方法などがあるが、シェッフエの方法がよく使われている。シェッフエの方法には、原法の他に芳賀変法、浦変法、中屋変法がある。

一対比較法は、2試料間の比較なので判断がしやすい、従って微妙な差も捉えやすいという利点があるが、試料数が多くなると組み合わせ数が多くなり労力的に大変になる。

3-3 特性描写

いくつかの評価項目をワンセットで評価することにより、試料の品質全体を総括的に捉え、かつ詳細に描写する方法である。

プロファイル法

プロファイル法の訓練を受けた3~10人のパネルが次のような評価を行う。1. 各特性毎に、その強さを評価する。2. フレーバーを感知する順序を判断する。3. 後味を評価する。4. アンプリチュード（全体の印象の強さ）を評価する。結果の解釈と解析は、円卓法による意見交換のすえ、パネルリーダーによりまとめられる。

QDA法

Quantitative Descriptive Analysisの略である。前述のプロファイル法を、より客観的データが得られるように改良したものである。主な特徴は次のようである。特性を表現する項目は、パネル訓練時点で定義され共有化される。スコアシートには評価項目が、予めフレーバーを感知する順番に、グラフ尺度（直線上に目盛りを付けた尺度）で与えられている。パネリストに、同一試料について数回ずつくり返し判断させることにより、パネリスト毎の能力評価も出来るようになっている。データは分散分析や主成分分析などの統計解析にかけられる。レーダーチャートにより品質特性が視覚的に描写される。

SD法

Semantic Differential法の略である。特に嗜好型官能検査で、品物のイメージを調べたい場合に用いられる。良い-悪い、重い-軽い、男性的-女性的、派手な-地味な などのような反対語を両端においた5~7段階尺度を10~30個提示し評定してもらう。各項目の平均値は、そのまま平均プロフィールとなるが、さらに因子分析や主成分分析により、各品物間の基本的なイメージ構造をさぐることもある。

4 統計的検定の考え方（2点識別法の場合）

2点識別法は、標準品と異常品の比較とか現行品と改良品の比較のように、客観的な順位が存在する場合に適用される。品質の良否や嗜好を調べる場合のように、客観的順位が存在しない場合には、2点嗜好法を用いなければならない。2点識別法と2点嗜好法は検定法に違いがある。2点識別法では、あらかじめ客観的に順位がついている試料に対して、検査員が正しく選べるかどうかを調べるので片側検定となる。これに対して2点嗜好法では、2つの試料間に客観的順位が存在しないので、両側検定となる。

4-1 2点識別法の検定原理

ある検査員がこのテストを1回行ったところ正解であった。この検査員は、識別出来たといえるだろうか。この場合、まぐれ当たりの確率が1/2あるので、識別出来たのかまぐれ当たりしたのか分からない。次に、このテストをn回くり返して、k回正解をした場合について考える。まぐれ当たりが起こる確率は、nに対するkの値で決まる。もし、この場合のまぐれ当たりの確率が非常に小さい値（通常5%以下）のときには、めったに起こらないケースと考え、まぐれ当たりではなく、識別されていると判定される。逆に、大きい値のときには、まぐれ当たりかもしれないので、識別されているとはいえないと判定される。

2つの試料を比較してどちらが正解かを当てる場合、n回のくり返しでk回の正解を得る確率P(k)は、次のような2項分布の一般式に従う。pは正解率（識別能力）である。

$$P(k) = {}_n C_k p^k (1-p)^{n-k}$$

検定をするために、まず2つの試料の差を識別出来ないという帰無仮説をたてる。

帰無仮説 $H_0 ; p = 1/2$

すなわち、pが1/2ということは、「判定結果は、まぐれあたりである」という仮説である。

この場合、P(k)は次の式になる。

$$P(k) = {}_n C_k (1/2)^n$$

この帰無仮説のもとで、n回のくり返しで、k回以上正解（この場合、まぐれ当たり）する確率は、次のような2項分布の確率から計算できる。

$$P = P(k) + P(k+1) + \dots + P(n) = \sum_{x=k}^n {}_n C_x \cdot (1/2)^n$$

この式より求められるPの値が、あらかじめ定められた確率α（通常α=0.05）以下であれば、危険率αで帰無仮説を棄却し、2つの試料間に有意差があるとみなす。

4-2 実施方法

2種類の試料A, Bを1組とする提示用試料をn組、すなわち

$$n = (\text{検査員数}) \times (\text{一人の検査員のくり返し数})$$

の数だけ用意する。そして、検査員に対して提示試料のセットをくり返しの数だけランダムな順序で提示し、正解を指摘させる。この時、A→B, B→Aの2とおりの比較があるため、順序効果を消すため、両方の比較順を同じ数にするのが望ましい。データを集計して、判定回数nに対する正解数kをもとめる。

2点識別法のための検定表（表3. 参照）でkと限界値を見比べれば、ただちに有意差があるか

表3 2点識別法の検定表

(危険率5%の場合)	
総判定数	限界値
5	5
6	6
7	7
8	7
9	8
10	9
11	9
12	10
13	10
14	11
15	12
20	15
30	20
40	26
50	32

kが限界値以上であれば有意

どうか判定できる。二点識別法の場合、判定回数 n は最低5回必要である。 $\alpha = 0.05$ の危険率で検定する場合、 $n = 5$ では $k = 5$ 、すなわち全問正解でないとは有意差があるといえない。 $n = 6$ で $k = 6$ 、 $n = 7$ で $k = 7$ 、 $n = 8$ でやっと $k = 7$ となる。 $n = 10$ でも $k = 9$ である。従って、2点識別法で判定しようとするときは、 n がかなり大きくなるように計画しなければならない。

なお、2点嗜好法の場合には、両側検定による検定表を用いて判定すればよい。

5 官能検査の適用分野に対する主な手法

企業において官能検査が適用される場面は非常に多くある。まず、新製品開発分野では、商品コンセプトをもとに、試作品→(評価)→1次改良品→(評価)→2次改良品→ というステップで、

次々とさまざまな評価がされ最後に完成品が出来上がる。このステップでは、既存製品と試作品の比較、競合する他社製品と試作品の比較、試作品の品質特性描写、試作品の保存性試験(シエルフライフテスト)などがよくおこなわれる。また、製品の品質改善分野でも、製造工程の改善、製造コスト削減、新原料の選択などに伴う品質の変化をチェックすることが必要となる。改善品がねらいどおりの品質に仕上がっているかどうかを調べるため、従来品と改善品との比較評価がおこなわれる。さらに、工場生産分野においては、原料受け入れ検査、工程検査、出荷検査などが定常的におこなわれている。また、品評会、鑑評会での品質評価といったケースもある。そして、新製品や改良品にたいする顧客の反応を調べるために、市場テストもよくおこなわれる。表4に企業活動

表4 官能検査の適用分野・問題の形式と主な手法

適用分野	問題の形式	手法
A. 新製品開発 まったく初めての製品 他社または既存製品の変形	<ul style="list-style-type: none"> 試作品の品質特性描写 標準品と比較して、試作品の検討を行う。 標準品に比べて、嗜好性が向上しているか。 	<ul style="list-style-type: none"> プロファイル法、QDA法 2点、3点識別法 カテゴリー尺度法、採点法 2点、3点嗜好法、順位法 嗜好尺度法
B. 製品の品質改善 工程改善 コスト削減 新原料の選択	<ul style="list-style-type: none"> 試作品が標準品に比べて差があるか。 試作品が標準品に比べて好まれているか。 	<ul style="list-style-type: none"> 2点、3点識別法 カテゴリー尺度法、採点法 2点、3点嗜好法、順位法 嗜好尺度法
C. 製品品質の維持	<ul style="list-style-type: none"> 標準品に比べて差があるか。 	<ul style="list-style-type: none"> 2点、3点識別法 カテゴリー尺度法、採点法
D. 製品の保存性	<ul style="list-style-type: none"> 保存品が標準品に比べて差があるか。 品質変化の程度の尺度化と特性描写 	<ul style="list-style-type: none"> 2点、3点識別法 カテゴリー尺度法、採点法 カテゴリー尺度法、QDA法
E. 品評会、鑑評会	<ul style="list-style-type: none"> 出品試料の格付け、採点 	<ul style="list-style-type: none"> 格付け法、採点法 カテゴリー尺度法
F. 新製品または改良品の 市場テスト	<ul style="list-style-type: none"> 評価したいサンプルが1点の場合 <ul style="list-style-type: none"> ア. 対象品がない場合 イ. 対象品と比較する場合 評価したいサンプルが数点の場合 	<ul style="list-style-type: none"> 嗜好尺度法、嗜好意欲尺度法 SD法 2点、3点嗜好法、嗜好尺度法、嗜好意欲尺度法、SD法 順位法、一対比較法 嗜好尺度法、嗜好意欲尺度法、SD法

において、官能検査が適用される分野及び問題の形式と、それに対応して適用される主な官能検査手法を示した。

6 おわりに

以上、官能検査手法について概説した。実際に手法を適用される場合には、後記の参考文献が役立つであろう。特に、文献2には、各手法の統計原理と計算手順が分かりやすく、かつ詳しく説明されている。得られたデータについて、有意差検定をするには、検定表（早見表）が整備されているので簡単に検定出来る。また、特性描写法など複雑な解析を要するものについては、コンピューターソフトとして、JUSE-QCAS, JUMP, STATVIEWなどを利用すれば容易に解析出来る。

目的に応じた官能検査手法を、気軽に使いこなされることを期待いたします。

[参考文献]

1. 官能検査入門 : 佐藤 信, 日科技連
 2. 統計的官能検査法 : 佐藤 信, 日科技連
 3. 新版官能検査ハンドブック : 日科技連官能検査委員会編, 日科技連
 4. センソリーエバリュエーション-官能検査へのいざない- : 増山英太郎・小林茂雄, 垣内出版
 5. おいしさを測る : 古川秀子, 幸書房
 6. 朝倉康夫 : 官能評価データの解析, 品質, 15 : 76 - 81 (1985)
-