

シリーズ

研究の動向 8

官能評価法—順位法について—

市 原 茂

1. はじめに

食品の領域でよく用いられる官能評価における測定の水準は、名義尺度 (nominal scale)、順序尺度 (ordinal scale)、間隔尺度 (interval scale)、比率尺度 (ratio scale) の4つの水準からなるとされている¹⁾。今回問題とするのは順序尺度で、順序尺度というのは測定値の大小の順番は意味を持つが、測定値同士の値の差や比は意味を持たない、あるいは、それらを求めてはいけないう尺度をいう。得られた官能データが順序尺度の水準である場合には、標本の平均値や標準偏差を求めることは不適切で、求めることができるのは、中央値や四分位偏差などである。

例えば5つの試料を好きな順に並べたり、味が辛い順に並べたときの順番のデータは、そのまま順序尺度データになる。順序尺度の問題を考えるとときに浮上するのは、一見間隔尺度を満たしているように見えるけれども、厳密には間隔尺度を満たしていないデータの扱いの問題である。試料の好き嫌いを評定する方法として、非常に好きを5点、非常に嫌いを1点として、あとは好みに応じて1点から5点の間のいずれかの点をつける採点法 (scoring method) という方法がある。採点法は、評定尺度法 (rating method) ともいわれ、態度尺度構成法の一つであるリッカート法 (Likert method)²⁾ も同様の方法である。採点法の場合、数値間の心理的間隔が等しくなるように判断して下さいと評定者 (パネル) に要求する。そして、それを根拠にして、得られたデータを間隔尺度とみなし、データをなんら加工せずに平均値や標準偏差を求めている。しかし、このような教示だけで間隔尺度がはたして保証

されるのかどうかは、疑問である。同じ問題は、SD法についてもいえる。例えばSD法では、ある物を見て、非常に美しいと感じたら7点、非常に汚いと感じたら1点というように、「美しい—汚い」という反対の意味を持つ形容詞の軸上で評定を行わせるが、SD法の場合も、多くの場合、得られた生のデータをそのまま用いて平均値を求めたり、因子分析などを行っている。これも厳密に言えば、問題がないわけではない。

採点法によって得られたデータの尺度の水準が間隔尺度を満たしていない場合に、解決方法として二つの方法がある。一つは、データを加工して、尺度の水準を間隔尺度の水準まで高めるもので、一例として、サーストンの系列範疇法 (method of successive categories)³⁾ を用いて、それぞれの評点の間隔を心理的に等間隔になるように修正する方法がある。系列範疇法は、ある刺激に対する反応の分布が、心理学的連続体の上で正規分布をなすという仮説の下に行われる。もちろん、この仮説がくずれたら、この手法は成り立たないが。他の一つは、間隔尺度のデータとしての扱いをせず、評定値の大きさの順に順位値を割り振って、順序尺度のデータとして処理するという方法である。

評点法ではなく、はじめから順位を答えてもらう順位法を用いた場合にも、サーストンの正規化順位法 (normalized rank approach)³⁾ のように、順序尺度データを間隔尺度に高めて、試料の尺度値 (間隔尺度) を求める方法や、あくまでも順位尺度のデータとしての扱いを続け、試料の順位値の相関を求めたり (スピアマンの順位相関など)、差の検定をしたりする (フリードマンの順位検定など) 方法がある。順位法については、ISO (International Organization for Standardization: 国際標準化機構) が定めた規格 (ISO8587: 2006 (E)) として、スピアマンの順位相関、ページ検定、フリードマンの順位検定が取り上げられている。サーストンの正規化順位法やケンドールの順位相関、ウィルコクソンの順位和検定などは取り上げられていない。そこで、今回は、フリードマンの順位検定を中心に解

Shigeru ICHIHARA 首都大学東京

著者紹介【略歴】東京都立大学人文学部助手、中京大学文学部講師、同助教授、東京都立大学人文学部助教授、同教授を経て、首都大学東京人文科学研究科教授、文学博士、日本官能評価学会会長、日本基礎心理学会理事、日本人間工学会感性情報処理・官能評価部会長、ISO/TC34/SC12 国内対策委員。【専門分野】心理学 (知覚心理学)、官能評価。

表 1. 順位法：完備型ブロック計画データ例

評価者	試料 1	試料 2	試料 3	試料 4	試料 5
1	2	4	5	3	1
2	4	5	3	1	2
3	1	4	5	3	2
4	1	2	5	3	4
5	1	5	2	3	4
6	2	3	4	5	1
7	4	5	3	1	2
8	2	3	5	4	1
9	1	3	4	5	2
10	1	2	5	3	4
順位和	19	36	41	31	23

説することとする。

2. フリードマンの順位検定

順位法には、完備型ブロック計画とつりあい不完備型ブロック計画がある。完備型ブロック計画というのは、一人の評価者がすべての試料の順位づけを行うもので、その際、評定者は、試料の数に応じて1位から暫時順位をつけていき、同順は認められない。試料の呈示順序はランダム順である。つり合い不完備型ブロック計画というのは、一人の評価者がすべての試料を評価するのではなく、評価者全体の結果を加えると各試料の評定回数が均等になるように配分された特定の組み合わせの試料だけを行うものをいう。

<完備型ブロック計画>

表1は、10名の評定者に5つの試料に1—5の順位値をつけさせたときのデータの例である。試料間の順位値に差があるかどうかを検定するために、まず、それぞれの資料ごとに順位和 R_j を求める。次に、(1)式より、 F_{test} 値を求め、この値の高低により、試料間の順位値に差があるかどうかを検定することになる。この場合、帰無仮説 H_0 と対立仮説 H_1 は、以下のようになる。

H_0 ：試料間の順位値に差がない。

H_1 ：試料間の順位値に差がある。

$$F_{\text{test}} = 12(R_1^2 + \cdots + R_p^2) / \{jp(p+1)\} - 3j(p+1) \quad (1)$$

(但し、 R_j ：試料 j の順位和。表1の場合、 R_j^2 の総和は、 $19^2 + 36^2 + 41^2 + 31^2 + 23^2 = 4828$ 。 j ：評価者数(10)、 p ：試料数(5))

求める F_{test} 値は、

$$F_{\text{test}} = 12 \times 4828 / (10 \times 5 \times 6) - 3 \times 10 \times 6 = 13.12 \quad (3)$$

となる。

ISO 8587: 2006 の Table 4 に試料の数(5)と評定者の数(10)による5%水準及び1%水準の棄却限界値が示されているが、これによると1%水準で棄却限界値は12.38、5%水準では9.25である。求めた13.12は、そのいずれよりも高い値を取るため、表1の例では、1%水準で試料間の順位値に差があったといえる。

なお、試料数や評定者数が多く、Table 4 に載っていない場合は、得られた F_{test} 値を自由度 $p-1$ の χ^2 値とみなし、 χ^2 検定を行う。

試料間に有意な差があった場合に、次になすべきことはどの試料とどの試料の間に有意な差があったのか、下位検定をすることである。5%水準あるいは1%水準で有意となる順位和の最小の差 Least Significant Difference (LSD) を求める。LSD を求める際に、危険率 α を実験全体に適用すれば、それぞれの試料のペアでは、危険率 α' は、 $\alpha' = 2\alpha/p(p-1)$ となる。一方、危険率 α を実験全体に適用するのではなく、個々のペアに適用するのであれば、危険率は α のままで変わらないことになるが、多くの場合、前者が適用されるし、その方が、妥当なことが多い。後者が適用されるのは、個々のペアが独立な場合だけであり、通常の事態では考えにくい。

今回の例では、 $p=5$ 、 $\alpha=0.05$ であるから、 $\alpha' = 0.005$ になり、これに対応する z 値は、2.81 になる。

LSD は、(3)式で与えられるので、

$$LSD = z\sqrt{jp(p+1)/6} \quad (3)$$

求める LSD は、19.87 となる。表1より、順位和の差が19.87よりも大きいのは、試料1と3の間の差($41-19=22$)だけである。よって、試料1と試料3の間に5%水準で有意差があったといえる。

<つり合い不完備型ブロック計画>

表2は、つりあい不完備型ブロック計画で10名の評価者に5つの試料を評価してもらったときのデータの一覧表である。それぞれの評価者は、5つの試料の内3つの試料にだけ順位をつけ、全体として、それぞれの試料が評価される数は、試料間で等しくなっている。

まず、試料間の順位値に差があるかどうかを検定するために、(4)式より、 F_{test} 値を求める。

$$F_{\text{test}} = 12(R_1^2 + \cdots + R_p^2) / \{rgp(k+1)\} - 3m^2(k+1)/g \quad (4)$$

研究の動向

表2. 順位法：つりあい不完備型ブロック計画データ例

評価者	試料 1	試料 2	試料 3	試料 4	試料 5
1	1	2	3		
2	1	2		3	
3	2	3			1
4	1		2	3	
5	2		3		1
6	1			3	2
7		1	3	2	
8		2	3		1
9		3		2	1
10			1	3	2
順位和	8	13	15	16	8

(但し, R_j : 試料 j の順位和. 表 2 の場合, R_j^2 の総和は, $8^2 + 13^2 + 15^2 + 16^2 + 8^2 = 778$. j : 評価者数 (10), p : 試料数 (5), r : 繰り返し数 (1), k : 各評価者が評価する試料の数 (3), n : 各試料が評価される回数 (6), g : 1 対の試料が一緒に評価される回数 (3))

求める F_{test} 値は,

$$12 \times 778 / (1 \times 3 \times 5 \times 4) - 3 \times 1 \times 6^2 \times 4 / 3 = 11.6 \quad (5)$$

11.6 は, 5%水準の棄却限界値 9.25 よりも大きく, 1%水準の棄却限界値 12.38 よりも小さいため, 表 2 の例では, 5%水準で試料間の順位値に差があったといえる.

次に下位検定を行う. LSD は, (6)式で与えられる. 今回も $p=5$, $\alpha=0.05$ であるから, $\alpha'=0.005$ になり, これに対応する z 値は, 2.81 になる.

$$\text{LSD} = z \sqrt{r(k+1)(nk-n+g)/6} \quad (6)$$

(6)式より求める LSD は, 11.6 となった. 表 2 より,

順位和の差は, 最大でも試料 1 と 4, あるいは, 試料 4 と 5 の間の 8 であることから, 有意差のあるペアはなかったことになる.

以上がフリードマンの順位法の概要である.

3. おわりに

すでに述べたように ISO8587: 2006(E) では, 統計法としては, 今回紹介したフリードマンの順位検定の他に, スピアマンの順位相関とページ検定が取り上げられている. これらは, いずれもノンパラメトリック検定である.

今回, サーストンの正規化順位法が取り上げられなかったというのは, 一つの時代の流れを映し出しているような気がする. 感覚の尺度の次元に正規分布を仮定することで尺度の水準を上げるというのが, 一昔前は常套手段だったような気がするのだが, 今日では, 無理な仮定は置かず, 順序尺度は順序尺度のままで扱うという流れになってきているのかもしれない. 多変量解析でも, いわゆる非計量的多次元解析法などが成果をあげているように.

また, ISO8587: 2006(E) では, 上記の統計法以外にも, 順位法を実施する際の検査環境や検査手続きなど, 詳細にわたる指示が記されている. これらを遵守することにより, 実りある成果が期待できる.

引用文献

- 1) Stevens, S. S.: On the Theory of Scales of Measurement, *Science*, **103**, 670-680 (1946)
- 2) Likert, R.: A Technique for the Measurement of Attitudes, *Archives of Psychology*, No. 140, 1-55 (1932)
- 3) J. P. ギルフォード (秋重義治監訳): 『精神測定法』, 培風館, 東京 (1959)