

製品開発に役立つ感性・官能評価データ解析

— R を利用して —

まえがき

人間の五感に深い満足感を与え、それを所有していることだけで満たされた気分になるようなものづくりをするためには、何が必要なのだろうか。

ものの良し悪しやその特性は、機械が判断するのではなく、人間が、その感覚で評価してこそ、正しい結論にたどり着くという考えの下、感覚によるものの評価を推し進め、その手法を確立してきたのが、官能評価という学問である。もちろん、ものによっては、感覚による評価では、最適な結論に到達できないという領域もあるかもしれない。しかし、最終的に人間が使用する製品を作る限りにおいては、人間の特性を無視してものづくりをすることは、考えにくい。使用する人間の特性に適合しなければ、その製品は人間に満足をもたらすことはできず、それだけではなく、人間に不都合な結果をもたらすかもしれないのである。感覚で評価する場合、物理的な測定器で測る物理的な値とは異なり、客観的な測定単位があるわけではないので、どのようなものさしを用いて、感覚の大きさやものの特性を表現するのが適切なのかという問題が浮いてまわるのも事実である。

このむずかしい問題に答えるために、官能評価では様々な手法が提案され、それらの手法のいくつかは、JIS や ISO などにより、基準化されている。

これらの手法を適切に用いることにより、ものの官能特性は正確に表現されることになる。そして、このようにして測定された官能データを蓄積し、それらを適切に整理統合すれば、人間の感性に合致する新たな感性製品の開発にも結び付くことになると思われる。

なお、本書は二部で構成されており、第1部解説編では、様々な感性・官能手法を紹介・解説し、第2部解析編では、無料のプログラムソフト R を利用して、1部で紹介した手法を用いて測定したデータの解析プログラムを紹介している。

ものづくりの現場にいる方々や、ものの官能特性を評価する部署において、官能評価手法を取り入れている方々、あるいは、官能評価手法を学んで将来の自分の進路に役立てたいと考えている方たちの一助になりたいと考えている。

なお、本書で紹介する解析プログラムに関しては、HP を通して利用できるようにした。以下の HP を参照願いたい (<http://www.j-sems.com>)。このプログラムを利用したユーザー同士が情報を提供しあうことにより、さらに使い勝手の良い利用環境が整うことも期待している。これらを通して、人間の感性を考慮した感性製品が、開発される糸口としたいと考えている。

2015年11月

株式会社 メディア・アイ

目次

第1部 解説編

1. 官能評価とは	1
1-1. 官能評価におけるパネル	1
1-2. 官能評価のデータの性質	1
1-3. 官能評価の手法	3
2. 官能評価手法	5
2-1. 識別試験法	5
1) 2点試験法	5
(1) 2点識別法	5
(2) 2点嗜好法	8
2) 3点試験法	10
3) 1対2点試験法	12
4) 配偶法	12
2-2. 順位法	15
1) スピアマンの順位相関係数	15
2) ケンドールの順位相関係数	17
3) ページの検定	19
4) ケンドールの一致性係数	21
5) フリードマンの順位検定	24
(1) 完備型ブロック計画	24
(2) つり合い不完備型ブロック計画	27
6) サイン検定	29
7) ウィルコクソンの順位和検定	31
8) クラスカル-ウォリスのH検定	35
9) 正規化順位法	39
2-3. 一対比較法	40
1) 一意性の係数	40
2) 一致性の係数	42
3) ブラッドレイの一対比較法	45
4) シェフェの一対比較法	50
(1) 原法	50
(2) 浦の変法(変形1)	58
(3) 芳賀の変法(変形2)	63
(4) 中屋の変法	67
5) サーストンの一対比較法	72
2-4. 格付け法	78
2-5. CATA法(コ克蘭のQ検定、対応分析)	81

2-6. 採点法	83
1) 平均値の差の検定 (t 検定)	83
(1) 対応のあるデータの場合	83
(2) 対応のないデータで、等分散の場合	84
(3) 対応のないデータで、等分散でない場合	86
2) 分散分析	88
(1) 一元配置の分散分析	88
(2) 二元配置の分散分析	90
(2-1) 繰り返しがない場合	90
(2-2) 繰り返しがある場合	93
(2-2-1) 両要因とも母数モデルの場合	93
(2-2-2) 混合モデル	96
(2-3) つり合い不完備型ブロック計画 (BIBD)	98
3) 積率相関	104
4) 回帰分析	105
(1) 単回帰分析	105
(2) 重回帰分析	107
5) 多変量解析 (判別分析、因子分析、主成分分析)	110
(1) 判別分析	110
(2) 主成分分析	112
(3) 因子分析	115
(4) コンジョイント分析	119
2-7. SD 法	123
2-8. QDA® 法 (線尺度あるいは VAS(Visual Analog Scale))	124
2-9. マグニチュード推定法	125
2-10. 時系列評価法	128
1) TI (Time Intensity)	128
2) TDS (Temporal Dominance of Sensation)	130
3) TCATA (Temporal Check-ALL-That-Apply)	133
2-11. 心理物理的測定法	136
1) 調整法	136
2) 極限法	136
3) 恒常法	137
4) 適応的方法	139
(1) 上下法	139
(2) QUEST	140
5) 信号検出理論	142
3. 官能評価手法についての問答集	147

第 2 部 解析編：R による分析

4. 識別試験法	152
4-1. 2点試験法	152
4-2. 2点嗜好法	152
4-3. 3点試験法	153
4-4. 1対2点試験法	153
4-5. 配偶法	154
1) 配偶法1 < t 個同士のマッチングで繰り返しのない場合 >	154
2) 配偶法2 < t 個同士のマッチングで繰り返しのある場合 >	156
3) 配偶法3 < t 個と (t + 1) または (t + 2) 個のマッチングの場合 >	157
5. 順位法	159
5-1. スピアマンの順位相関係数	159
5-2. ケンドールの順位相関係数	160
5-3. ページの検定	161
5-4. ケンドールの順位の一貫性係数	162
5-5. フリードマンの順位検定	163
1) 完備型実験計画	163
2) つりあい不完備型実験計画	166
5-6. サイン検定	169
5-7. ウィルコクソン順位和検定	169
5-8. クラスカル-ウォリスのH検定	170
6. 一対比較法	171
6-1. 一意性の係数	171
6-2. 一貫性の係数	174
6-3. ブラッドレイの一対比較法	176
6-4. シェフェの一対比較法	178
1) シェフェの方法(原法)	178
2) 浦の変法(変形1)	185
3) 芳賀の変法(変形2)	191
4) 中屋の変法(変形3)	196
6-5. サーストンの一対比較法	202
7. 格付け法	206
7-1. χ^2 検定	206
7-2. フィッシャーの直接確率法	207
8. CATA 法 (対応分析)	208
9. 採点法	211
9-1. 2群の平均値の差の検定	211
1) 対応のあるデータの平均値の差の検定	211
2) 対応のないデータの平均値の差の検定 (分散が等しい場合)	212

3) 対応のないデータの平均値の差の検定 (分散が等しくない場合)	213
9-2. 分散分析	214
1) 一元配置の分散分析	214
2) 二元配置分散分析 (繰り返しなし)	216
3) 二元配置分散分析 (繰り返しあり)	217
4) 二元配置分散分析 (つり合い不完備型ブロック計画: BIBD)	218
9-3. 積率相関係数	222
9-4. 単回帰分析	223
1) 単回帰分析 1	223
2) 単回帰分析 2 (マグニチュード推定法のデータ分析)	224
10. 多変量解析	227
10-1. 重回帰分析	227
10-2. 判別分析	228
10-3. 主成分分析	233
10-4. 因子分析	237
10-5. コンジョイント分析	242
11. TDS	247
12. 数表	251
13. 著作権・商標・ライセンスに関して	257
14. 引用文献	258
15. 謝辞	260
16. 著者紹介	260

第 1 部 解説編

1. 官能評価とは

官能評価は、製品に対する人の五感の反応特性を測定し、分析し、解釈する学問であり、利用者の立場に立ったものづくりをしたいと考える人々にとっては、非常に有用な学問といえる。

ものづくりは、それを使用する人間の特性を抜きにしては成り立たないものといえる。実際、様々な現場でものづくりに励んでいる人は、官能評価の手法を利用するかどうかは別として、自分の関わっている製品が、世の中の人たちに、どのように評価され、受け入れられているのであろうかと思ひめぐらすのではないだろうか。製作者の感性と勘に頼ってものづくりを行う場合もあるであろうが、その決定に、客観的な根拠を与えてくれるのが、官能評価ともいえる。

また、官能評価は、製作者の狙った通りに、一般の消費者に評価されているのかどうかについても、データを提供してくれる。正しい手法で測定したものであれば、その評価データを次の製品の計画に利用することができると思われる。

実際、人の五感の反応特性の測定手法には、様々なものがあり、結果の分析方法も、データの質によって様々な方法がある。

官能評価の専門用語と手法については、ISO に様々な規格が取り上げられているが、本論では、官能評価の手法を中心に、どのような感性・官能データに対して、どのような手法を用いたらよいかを説明し、さらに、データの分析方法について解説する。

1-1. 官能評価におけるパネル

官能評価のために選ばれた集団のことをパネルという。なお、個人をさす時は、パネリストという。

パネルには、大きく分けて分析型パネルと嗜好型パネルがある。

前者は、試料の属性を分析的に試験する時に用いるパネルで、一定以上の感覚能力を要求される。分析型評価では、通常、パネルの人数は少人数であるが、10名以上は必要とされている。

後者は、試料に対する消費者の嗜好を予測したい時に用いるパネルで、嗜好型パネルという。また、嗜好型パネルを用いた官能評価のことを嗜好型評価という。ここでは、目的とする消費者を代表するようなサンプリングが問題になり、特別の感覚能力は要求されない。通常、嗜好型パネルの人数は、分析型パネルよりも多人数になる。

この他に、官能評価分析に関して豊富な経験を持ち、試料についても深い知識を持っているパネルのことを専門評価者という。専門評価者は、製品の変動する効果の評価や予測などを行うことができる。

また、後述の QDA 法で用いられる専門的なパネルのことを QDA パネルという。

1-2. 官能評価のデータの性質

官能評価では、人が製品を見たり、聞いたり、触ったり、味わったり、あるいは、そのにおいを嗅いだ時に感じる主観的な印象を、感覚データとして測定し科学的に分析する学問であるが、人が感じる主観的な印象をどのように表現し、測定し、官能データとするかは、とてもむずかしい課題といえる。官能評価では、感覚印象を様々な手法を用いて数値化（感覚印象に数字を付与）し、統計処理を行うのであるが、付与された数字が、実際の感覚印象を忠実に反映しているのかという点については、もう一度深く考える必要がある。実際、官能評価で扱われている官能データは、評価者の主観的な印象そのものではなく、あくまでも感覚印象を数字で置き換えたものであるため、その置き換えた数字が、感覚印象を忠実に反映しているという保証はないのである。そのような困難な問題を抱えつつも、その中で最適な解を見出すために、私たちは、現在

ある感性・官能手法の中で、当該の問題解決にとって最適な手法を選択しなくてはならない。

感性・官能手法を選択する際に考慮する必要があるのは、官能評価データの尺度の水準である。

官能評価データの尺度の水準には様々な段階があり、われわれは、問題とする現象のいろいろな側面を考慮し、どの水準の尺度を求めたらよいかを決定し、最適の手法を選択しなくてはならない。選んだ尺度の水準の適否によって、また、採用する官能評価手法の適否によって、研究の成果が決まるからである。

以下に、官能評価データの尺度水準について解説する。

<官能評価データの尺度水準>

心理学者のスティーブンスによれば、尺度には、以下の表のように、(1) 名義尺度 (2) 順序尺度 (3) 間隔尺度 (4) 比率尺度 の4つの段階があり、それぞれの尺度には、許される演算と、許されない演算があるという。官能評価にも、この考えをあてはめることができ、官能評価データの尺度の水準により、適用する官能手法の候補が決まることになる。そして、それらの候補には、許容される尺度の水準に加えて、様々な特性があることから、それらを総合的に考慮して、最適な手法を選択しなくてはならない。

表 1.2.1 尺度の水準

尺度	尺度の説明	データ、統計量、演算の例	該当する官能手法
名義尺度	分類することによって得られる尺度.	質的データ、男女の分類など. 度数、最頻値、 χ^2 検定	格付け法
順序尺度	順位がつけられる尺度. 各尺度値間の差は意味を持たない.	1番・2番、あるいは、優・良・可などの順位データ。 名義尺度に許された演算に加えて、中央値や順位相関係数など.	順位法 格付け法
間隔尺度	順位がつくのと同時に、各尺度値間の差も意味を持つ.	摂氏、華氏などの温度（絶対的な零点がない（絶対温度ではない）ので、温度の差は表現できるが、比は表現できない）。 順序尺度に許された演算に加えて、平均値や標準偏差、積率相関係数など.	採点法
比率尺度	各尺度置間の比が意味を持つ尺度.	絶対温度、長さ、重量など。絶対的な零点があるため、尺度間の比を取ることができる。 間隔尺度に許された演算に加えて、比、幾何平均、調和平均、RMS（二乗平均平方根）、変動係数など.	マグニチュード推定法

1-3. 官能評価の手法

官能評価で用いられる手法には、目的に応じて様々なものがある。また、1-2で解説したように、感性・官能データの尺度の水準によって、用いられる手法も限定されてくる。

それぞれの官能評価手法の特徴を、表 1.3.1 に示す。

表 1.3.1 官能評価手法とその特徴

官能評価手法	手法の説明	データの尺度水準
2点識別法	パネルの識別能力を決定する方法	名義尺度
2点嗜好法	パネルの試料に対する嗜好や良否を決定する方法	名義尺度
3点試験法	パネルの識別能力を決定する方法	名義尺度
1対2点試験法	パネルの識別能力を決定する方法	名義尺度
配偶法	パネルの識別能力を決定する方法	名義尺度
スピアマンの順位相関係数	2変数間の順位の関連性の強さの指標	順序尺度
ケンドールの順位相関係数	2変数間の順位の関連性の強さの指標	順序尺度
ページの検定	評価されたサンプルの順位が、想定した順位通りであったかどうかを検定	順序尺度
ケンドールの一致性係数	複数のパネル間の順位の評定結果にどの程度の共通性があるかを示す指標	順序尺度
フリードマンの順位検定	試料間で順位づけされた順位に差があるかどうかを検定	順序尺度
サイン検定	評価者が下した2つの試料の順位の判定に差があるかどうかを検定	順序尺度
ウィルコクソンの順位和検定	m個のAとn個のBをこみにして順位をつけたとき、AとBの2組の順位付けに差があるかどうかを検定	順序尺度
クラスカル-ウォリスのH検定	ウィルコクソンの順位和検定と同様であるが、3種類以上の順位の差を検定するように拡張したもの。	順序尺度
正規化順位法	試料を順位づけたデータを、間隔尺度の評価値に変換する手法	順序尺度⇒ 間隔尺度
一意性の係数	評定者個人の対比較判断に統一性（一意性）がどれくらいあるかの測度	順序尺度
一致性の係数	n人の対比較の判断結果の一致度	順序尺度
ブラッドレイの対比較法	どちらが良いかの対比較データから判定比を逐次近似により推定する	順序尺度
シェフェの対比較法	対比較判断の程度を数量で評定する。評点をつける対比較法	間隔尺度

サーストンの一対比較法	サーストンの比較判断の法則に基づき、大小判断の一対比較データから間隔尺度を構成する	順序尺度→ 間隔尺度
格付け法	試料を特級、1級、2級へ分類したりする方法。クロス表による χ^2 検定を行なうのが一般的	順序尺度
採点法	数値尺度を使って、試料の特性や好ましさに対して評点を与える。分散分析、因子分析をはじめ、様々なパラメトリックな統計解析が可能。	間隔尺度
SD法	試料に対する印象を両側に反対語をなす形容詞対を伴った多くの評定尺度を用いて評定する	間隔尺度
線尺度	試料の特性を表現する様々な言葉を尺度にして評価する際に、その心理的な大きさを直線の長さで表現させる方法。	間隔尺度
マグニチュード推定法	心理的な大きさ（官能量）を比率に基づいた数量で表現する	比率尺度
Time Intensity	知覚される感覚強度の時系列的变化を記録し、得られた関数形の特性を様々なパラメータを用いて記述する方法	間隔尺度
TDS	複数の感覚の時系列变化を同時に測定する方法	間隔尺度
調整法	心理量・感覚量を測定する心理物理的測定法の一つ。	
極限法	心理量・感覚量を測定する心理物理的測定法の一つ。	
恒常法	心理量・感覚量を測定する心理物理的測定法の一つ。	
適応的方法	心理量・感覚量を測定する心理物理的測定法の一つ。	
信号検出理論	心理量・感覚量を測定する心理物理的測定法の一つ。	

2. 官能評価手法

2-1. 識別試験法

識別試験法は、パネルの識別能力や試料に対する嗜好や良否を決定する方法であり、2点試験法、3点試験法、1対2点試験法、配偶法などがある。

2点試験法、3点試験法、1対2点試験法、配偶法は、主に、試料間の差の有無やパネリストの識別能力の有無を判定するのに用いられる。

2点試験法は、どちらが甘いとか、どちらが好ましいかというように、一次元の特性を比較することを要求する場合には、識別力が高い方法であるが、あらかじめ、特性が明確にできないような試料間の差異の識別には用いることはできない。その場合には、3点試験法、もしくは、1対2点試験法を用いる。

配偶法は、多種類の試料の識別を同時に行うときに用いる手法の一つである。

1) 2点試験法

2点識別法と2点嗜好法とがある。いずれも二項検定を行うが、2点識別法と2点嗜好法の違いは、前者が片側検定を用いるのに対し、後者は両側検定を用いる点にある。なお、二項検定とは、測定値が2つのカテゴリーのいずれか一方に分類される時に、それぞれのカテゴリーの母集団における比率が、帰無仮説における比率と異なるか否かを吟味する検定のことをいう。

(1) 2点識別法

あらかじめ客観的な順位がついている2種の試料間の差をパネリストが識別する能力を持つか否かを検定する際に用いる。2つの試料を同時、または、継時的に示し、どちらの試料の強度が強いかを判定させ、その正答率を測定する。一人のパネリストがn回識別試験を行う場合とn人のパネリストが1回ずつ識別試験をおこなう場合とがある。

片側検定を用いる。

<例>

パネリストにAとBの内、どちらの試料の強度が強いか(AかB)を答えてもらう。

帰無仮説：パネリストはAとBを識別する能力を持たない。

n回中、k回正答で、n-k回誤った時、
その確率は、

$$\sum_{x=k}^n P(x) = \sum_{x=k}^n {}_n C_x \left(\frac{1}{2}\right)^n = \sum_{x=k}^n \frac{n!}{x!(n-x)!} \left(\frac{1}{2}\right)^n \quad (2.1.1)$$

となる。

例えば、10回中9回正答し、1回誤答した場合。

(2.1.1) 式より、その確率は、 $p(10)+p(9)$ となる。

$$P(10) = \frac{10!}{10! \times 0!} \left(\frac{1}{2}\right)^{10} = 0.000977 \quad (2.1.2)$$

$$P(9) = \frac{10!}{9! \times 1!} \left(\frac{1}{2}\right)^{10} = 0.009766 \quad (2.1.3)$$

より、

$$P(9)+P(10)=0.009766+0.000977=0.010743$$

となり、5% (0.05) よりも小さな値になることから、危険率5%で、パネリストがAとBを識別できないとする帰無仮説は棄却され、パネリストは、AとBを識別できるといえる。

また、10回中、8回正答で、2回誤った時、

$$P(8) = \frac{10!}{8! \times 2!} \left(\frac{1}{2}\right)^{10} = 0.043945 \quad (2.1.4)$$

となることから、

$P(8)+P(9)+P(10)=0.043945+0.009766+0.000977=0.054688 (>0.05)$ となり、0.05 よりも大きな値になることから、5%水準で有意でない、つまり、帰無仮説を棄却できない：パネリストはAとBを識別できないことになる。

$n>100$ の時は、次式により、正規分布近似を行ない検定する。

$$u_0 = \frac{k - \frac{n}{2} - 0.5}{\sqrt{n \times \frac{1}{2} \times \frac{1}{2}}} \quad (2.1.5)$$

$u_0 \geq u_\alpha$ ならば、帰無仮説を棄却する。

$\alpha = 0.05$ の時、 $u_{0.05} = 1.64485$

$\alpha = 0.01$ の時、 $u_{0.01} = 2.32635$

<事例>

試料Aと試料Bの2種類の苦み成分量の異なるポテトチップを用意する。なお、ここでは、試料Bの方が苦み強いものとする。20人のパネルがそれぞれ1回ずつ両者を比較し、どちらが試料Bか(どちらが苦いか)を答える。試料Bの方が試料Aよりも苦いと答えたパネルは、20名中17名であった時、これらのパネルは、試料Aと試料Bの苦みを識別できたといえるか。

考え方:pの計算は、上記の方法により直接求めるか、あるいは、検定表(表2.1.2)により行う方法がある。今回は、 $n=20$, $k=17$ である。片側検定を行う。

表 2.1.1 2点識別法の検定表 (片側検定) ¹⁹⁾

n	危険率			n	危険率			n	危険率		
	5%	1%	0.1%		5%	1%	0.1%		5%	1%	0.1%
4以下				16	12	14	15	28	19	21	23
5	5			17	13	14	16	29	20	22	24
6	6			18	13	15	16	30	20	22	24
7	7	7		19	14	16	17	40	26	28	31
8	7	8		20	15	16	18	50	32	34	37
9	8	9		21	15	17	18	60	37	40	43
10	9	10	10	22	16	17	19	70	43	46	49
11	9	10	11	23	16	18	20	80	48	51	55
12	10	11	12	24	17	19	20	90	54	57	61
13	10	12	13	25	18	19	21	100	59	63	66
14	11	12	13	26	18	20	22				
15	12	13	14	27	19	20	22				

検定表によれば、n=20, 危険率5%の棄却限界値は15、n=20, 危険率1%の棄却限界値は16、n=20, 危険率0.1%の棄却限界値は18なので、危険率1%水準で検査品の方がコントロールよりも苦いといえる。

(2) 2点嗜好法

2種の試料に対する嗜好や良否に有意な差があるかどうかを検定する際に用いる。2つの試料（例えば、AとB）を同時、または、継時的に示し、どちらの試料を好むかを判定させ、その回答率を測定する。n人のパネリストが1回ずつ判定を行う。両側検定を用いる。

試行数（パネリストの人数）をn、Aの方を好むと回答した数とBの方を好むと回答した数の内、度数の多い方をkとする。

帰無仮説：パネリストのAとBに対する嗜好に差はない。

n人の内、k人だけAまたはBを選ぶ確率は $(x \neq k)$ 、

$$P(x) = {}_n C_x \left(\frac{1}{2}\right)^n \times 2 \quad (2.1.6)$$

n人の内、k人がAまたはBを選ぶ確率は、

$$\sum_{x=k}^n P(x) = \sum_{x=k}^n {}_n C_x \left(\frac{1}{2}\right)^n \times 2 \quad (2.1.7)$$

となる。

そこで、例えば、10人の内、9人がAまたはBを選ぶ確率は、 $\{P(10)+P(9)\} \times 2$ となる。

$$P(10) = \frac{10!}{10! \times 0!} \left(\frac{1}{2}\right)^{10} = 0.000977 \quad (2.1.8)$$

$$P(9) = \frac{10!}{9! \times 1!} \left(\frac{1}{2}\right)^{10} = 0.009766 \quad (2.1.9)$$

であることから、

$$\{P(9)+P(10)\} \times 2 = \{0.009766+0.000977\} \times 2 = 0.021494$$

と、0.05以下であることから、5%水準で有意となり、パネリストのAとBに対する嗜好に差はないとする帰無仮説を棄却する（嗜好に差がある）ことになる。

10人の内、8人がAまたはBを選ぶ場合には、その確率は、 $\{P(8)+P(9)+P(10)\} \times 2$ で求められ、

$$P(8) = \frac{10!}{8! \times 2!} \left(\frac{1}{2}\right)^{10} = 0.043945 \quad (2.1.10)$$

であることから、 $\{P(8)+P(9)+P(10)\} \times 2 = \{0.043945+0.009766+0.000977\} \times 2 = 0.109376 (>0.05)$ となる。よって、5%水準で有意でない（帰無仮説を棄却できない、つまり、AとBに対する嗜好に差はない）ことになる。

$n \leq 100$ の時は、上記のように直接確率を計算する方法の他に、検定表（表 2.1.2）を利用する方法もある。

表 2.1.2 2点嗜好法の検定表（両側検定）¹⁹⁾

n	危険率			n	危険率		
	5%	1%	0.1%		5%	1%	0.1%
6	—	—	—	22	17	18	19
7	7	—	—	23	17	19	20
8	8	8	—	24	18	19	21
9	8	9	—	25	18	20	21
10	9	10	—	26	19	20	22
11	10	11	11	27	20	21	23
12	10	11	12	28	20	22	23
13	11	12	13	29	21	22	24
14	12	13	14	30	21	23	25
15	12	13	14	40	27	29	31
16	13	14	15	50	33	35	37
17	13	15	16	60	39	40	43
18	14	15	17	70	44	47	50
19	15	16	17	80	50	52	56
20	15	17	18	90	55	58	61
21	16	17	19	100	60	63	67

$n > 100$ の時は、次式により、正規分布近似を行ない検定する。

$$u_0 = \frac{k - \frac{n}{2} - 0.5}{\sqrt{n \times \frac{1}{2} \times \frac{1}{2}}} \quad (2.1.11)$$

$u_0 \geq u_\alpha$ ならば、帰無仮説を棄却する。

$\alpha = 0.05$ の時、 $u_{0.025} = 1.95996$

$\alpha = 0.01$ の時、 $u_{0.005} = 2.57583$

<事例>

90 人の消費者パネルに軽自動車 A と B の後部座席の乗り心地の評価をしてもらったところ、そのうち 60 人のパネルが、A の方が乗り心地が良いと答えた。A の方が後部座席の乗り心地が良いと考えて良いであろうか。

<考え方>

2点嗜好法は両側検定である。検定表（表 2.1.4）より、 $n=90$ で危険率 5% の棄却限界値は、55、危険率 1% の棄却限界値は、58、危険率 0.1% の棄却限界値は、61、なので、危険率 1% で、A の方が後部座席の乗り心地が良いといえる。

⇒ $n > 100$ の場合は、すでに述べた式により正規分布近似を行ない検定する。

2) 3点試験法

3つの試料を1組にし、その内の2つは同じ試料、1つは異なる試料にする。パネリストは、この中の異なる試料を選ぶのが課題である。n回の測定の中で、2種の試料間に差があるか、あるいは、一人のパネリストに識別能力があるかを測定する際に用いる。二項検定を行う。

パネリストに A, A, B からなる3つの試料の内、異なるのはどれかを答えてもらう (B と答えれば、正解)。なお、試料の提示の仕方は、時間順序、あるいは、空間位置をランダムにする。なお、A, B, B からなる3つの試料を提示して、A を正解とする場合もある。

帰無仮説：パネリストは A と B を識別する能力を持たない。

n 回中、k 回だけ正答する確率は、

$$P(k) = \frac{n!}{k!(n-k)!} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{n-k} \quad (2.1.12)$$

n 回中、k 回正答する確率は、

$$P = \sum_{x=k}^n \frac{n!}{x!(n-x)!} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{n-x} \quad (2.1.13)$$

となる。

<計算例>

10 回中、8 回正解し、2 回間違えたとする。

その確率は、

$$P = P(10) + P(9) + P(8) = 0.003404 \quad (P < 0.01)$$

で求められる。

なお、

$$P(10) = (1/3)^{10} \times (2/3)^0 = 0.00001693$$

$$P(9) = 10 \times (1/3)^9 (2/3) = 0.000339$$

$$P(8) = 45 \times (1/3)^8 (2/3)^2 = 0.003048$$

n > 100 の時は、次式により、正規分布近似を行ない検定する。

$$u_0 = \left\{ k - \left(\frac{n}{3} \right) - 0.5 \right\} / \sqrt{n \times \left(\frac{1}{3} \right) \times \left(\frac{2}{3} \right)} \quad (2.1.14)$$

$u_0 > u_\alpha$ ならば、帰無仮説を棄却する。

$$\alpha = 0.05 \text{ の時、 } u_{0.05} = 1.64485$$

$$\alpha = 0.01 \text{ の時、 } u_{0.01} = 2.32635$$

<事例>

印刷方式を変えたパンフレットに、違いがあるかどうかを調べるために、従来方式で印刷したもの (A) と、新しい方式で印刷したもの (B) を、A を 2 回、B を 1 回、ランダムな順に見せて、異なるパンフレット (奇数試料) を指摘させるテストを 15 人のパネリストに行ったところ、12 人が正解した。A と B のパンフレッ

トの刷り上がりに違いがあるといえるか。

<考え方>

3点試験法は片側検定である。n ≤ 100 の時は、上記のように直接確率を計算する方法の他に、検定表を利用する方法もある。

検定表(表 2.1.3)より、n=15 で危険率 5% の棄却限界値は、9、危険率 1% の棄却限界値は、10、危険率 0.1% の棄却限界値は、12、なので、危険率 0.1% で、A と B のパンフレットの刷り上がりに違いがあるといえる。

表 2.1.3 3点試験法の検定表¹⁹⁾

(表中の数値は、各有意水準における最小正答数が示されている)

回答数	5%	1%	0.1%	回答数	5%	1%	0.1%
5	4	5	—	22	12	14	15
6	5	6	—	23	12	14	16
7	5	6	7	24	13	15	16
8	6	7	8	25	13	15	17
9	6	7	8	26	14	15	17
10	7	8	9	27	14	16	18
11	7	8	10	28	15	16	18
12	8	9	10	29	15	17	19
13	8	9	11	30	15	17	19
14	9	10	11	40	19	21	24
15	9	10	12	50	23	26	28
16	9	11	12	60	27	30	33
17	10	11	13	70	31	34	37
18	10	12	13	80	35	38	41
19	11	12	14	90	38	42	45
20	11	13	14	100	42	45	49
21	12	13	15				

3) 1対2点試験法

パネリストに、まず標準試料（見本）を与え、次に、標準試料と同じものと異なるものの1組を試料としてパネリストに示す。パネリストは、示された1組の試料対の中から、標準試料（見本）と同じ試料を選び出す。

どれくらい正答するかが問題になり、パネリストの識別能力を測定したい時に用いる。

検定法は、2点識別法と同じで、二項検定を行う。片側検定である。

<事例>

新しく開発したスキンケア(B)は従来品(A)の香りにわずかな変更を加えたものである。この新製品は、従来品と違いが感じられるかどうかを調べた。20人のパネリストを用いて、それぞれのパネリストに、まず従来品(A)を対照試料として提示し、次に、従来品(A)と新製品(B)を提示し、どちらが新製品か識別させた。

その結果、20名中、16名(k=16)が正解した。新製品(B)と従来品(A)には、違いがあるといえるか。

考え方：pの計算は、2点識別法の方法により直接求めるか、あるいは、2点識別法で用いた検定表により行う方法がある。今回は、n=20, k=16の片側検定を行う。

2点試験法の検定表によれば、n=20, 危険率5%の棄却限界値は15、n=20, 危険率1%の棄却限界値は16、n=20, 危険率0.1%の棄却限界値は18なので、k=16ということから、危険率1%水準で、新製品(B)と従来品(A)には、違いがあるといえる。

4) 配偶法

t個からなる試料の組みを2組作る。パネリストには、第1組の1つの試料と同じと思われる試料を第2組から選択するというを行ない、最終的に、t対の組み合わせを作る課題を行なわせる。t対の組み合わせの内、何対正解したかということで、パネリストの識別能力を測定することができる。

なお、1人のパネリストに1回のみ判定で繰り返しを行なわない場合と、繰り返しを行なう場合とがある。

また、試料の組みを作る際に、t個とt個というように同数の組みにして揃える場合と、t個と(t+1)個、あるいは、(t+2)個というように同数にしない場合とがある。

帰無仮説は、t個の試料が区別できないである。

t個の試料が区別できなくて、その結果ランダムに対をつくるとすると、t対のうちs対がちょうど正しく組み合わせられる確率は、

$$P(s) = \frac{1}{s!} \left\{ \frac{1}{0!} - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \dots + (-1)^{t-s} \frac{1}{(t-s)!} \right\} \quad (2.1.15)$$

で与えられる。

この式を用いて、s対だけでなく、s対以上が正しく組み合わせられる確率をすべて加算する。その確率が、 α （例えば、5%）以下であれば、有意水準 α で帰無仮説は棄却され、t個の試料の区別ができることになる。なお、この式は、tが4以上になると急速に収束するので、試料の数が4以上(t \geq 4)の場合、その確率は、tとは無関係に決まってしまう。

表 2.1.4 は、上記の確率が 5 % 以下であるような s が示されている。なお、この表において、 t (試料の数) は 4 以上で、 n は繰り返し数である。正しく組み合わせられた数、または、平均値が表の値 s 以上であれば、危険率 5 % で有意となる²⁰⁾。

表 2.1.4 配偶法のための検定表 (t 個と t 個の試料をマッチさせた時)²²⁾

試料の数 $t \geq 4$, 繰り返し数 n , 有意水準 $\alpha = 5\%$

正しくマッチできた個数またはその平均値が表の値 s 以上の時

n	$n \times s$	S									
1	4	4.00	6	11	1.83	11	18	1.64	20	29	1.45
2	6	3.00	7	13	1.86	12	19	1.58	25	34	1.36
3	7	2.33	8	14	1.75	13	20	1.54	30	40	1.33
4	9	2.25	9	15	1.67	14	21	1.50			
5	9	1.80	10	16	1.60	15	23	1.53			

1) t 個同士のマッチングで繰り返しのない場合

互いに相異なる t 種類 ($t \geq 4$) の試料を 2 組用意する。次に、一人のパネリストに、同種の試料が 2 組用意してあることを知らせ、各組から 1 個ずつ取り出して同種の試料の組み合わせを作らせる。

手順 1: 正しい組み合わせの数 s を数える。

手順 2: 表 2.1.4 より、繰り返し 1 回 ($n=1$) の時の s の値よりも、 s が大きな値をとったばあいには ($s \geq 4$ であれば)、5 % 水準で有意である。

<事例>

パネリストの色相の識別能力を判定するために、色相がわずかに異なる A,B,C,D,E,F の 6 種類の色票を 2 組作る。それぞれのパネリストに、同種の試料が 2 組用意してあることを知らせ、各組から 1 個ずつ取り出して同種の試料の組み合わせを作らせる。

表 2.1.5 データ例

第 1 組	A	B	C	D	E	F
第 2 組	A	B	D	C	E	F
正誤	○	○	×	×	○	○

判定: 正しい組み合わせの数、 s を数え、 $s \geq 4$ であれば、5 % 水準で有意である。

(2) t 個同士のマッチングで繰り返しのある場合

(例) 3 人のパネリストに繰り返して実施した場合。(一人のパネリストが 3 回繰り返す場合も可)

表 2.1.6 データ例

第 1 組	A	B	C	D	E	F	
第 2 組 (パネリスト 1)	A	E	B	D	C	F	$s=3$
第 2 組 (パネリスト 2)	B	A	C	D	E	F	$s=4$
第 2 組 (パネリスト 3)	A	B	C	D	F	E	$s=4$

手順 1: 正しい組み合わせの数 s を数える。

手順2：正しい組み合わせの数 s の平均 s_{ave} を求める。この例では、 $s_{ave}=(3+4+4)/3=3.67$

手順3：繰り返し n の時の s_{ave} が、検定表（表 2.1.7）の値以上であれば、5%水準で有意。この場合は、 s_{ave} が $n=3$ の時の値 2.33 よりも大きな値だったので、5%水準で有意。

3) t 個と $(t+1)$ または $(t+2)$ 個のマッチングの場合

1組は t 個、もう1組は $(t+1)$ または $(t+2)$ 個を用意する。パネリストには、同種の試料が用意してあることを知らせ、各組から1個ずつ取り出して同種の試料の組み合わせを作らせる。

なお、 t 個と u 個 ($u>t$) のマッチングを行ったときの $p(s)$ は、以下の式で求められる。

$$P(s) = \frac{t!}{s!u!} \left\{ \frac{(u-s)!}{0!(t-s)!} - \frac{(u-s-1)!}{1!(t-s-1)!} + \dots + (-1)^{t-s} \frac{(u-t)!}{(t-s)!0!} \right\} \quad (2.1.16)$$

この式を用いて、 s 対だけでなく、 s 対以上が正しく組み合わせられる確率をすべて加算する。その確率が、 α （例えば、5%）以下であれば、有意水準 α で帰無仮説は棄却され、 t 個の試料の区別ができることになる。

表 2.1.8 は、上記の確率が5%以下であるような s が示されている。なお、この表において、 t （試料の数）は4以上で、 n は繰り返し数である。正しく組み合わせられた数、または、平均値が表の値 s 以上であれば、危険率5%で有意となる¹⁹⁾。

表 2.1.7 実験者用記録用紙 兼 出力表（例）

第1組	A	B	C	D	E	
第2組	A	B	D	F	E	C
正誤	○	○	×	×	○	

手順1：正しい組み合わせの数 s を数える： $s = 3$

手順2：検定表（表 2.1.8）から $t = 5$ の時の $(t+1)$ 個とのマッチングの限界値を読み取る：この場合、限界値は3なので、5%水準で有意。

手順3：結果、このパネリストは試料の識別能力があるといえる。

表 2.1.8 配偶法のための検定表²²⁾

(t 個と $(t+1)$ または $(t+2)$ 個の試料を
マッチさせた時：有意水準 $\alpha = 5\%$)
正しい組み合わせの数 s が表 2.1.1 の
値以上であれば有意

t	$t+1$	$t+2$
3	3	3
4	3	3
5	3	3
6以上	4	3

2-2. 順位法

順位法は試料に対する嗜好や感覚の強さを順位づけする方法である。順位法には、1) 2つの試料間の関連の強さや2人のパネリスト同士の関連の強さを測定するもの、2) 試料の評定順位が想定した通りであったかを検定するもの、3) 複数のパネル間の順位の評定結果に共通性があるかを測定するもの、4) 試料の評定順位に有意な差があるのかを検定するもの、5) 試料の順位データから順位の尺度値を求めるものがある。

1)の方法には、「スピアマンの順位相関係数」、「ケンドールの順位相関係数」があり、2)の方法には、「ページの検定」があり、3)の方法には、「ケンドールの一致性係数」があり、4)の方法には、「フリードマンの順位検定」、「サイン検定」、「ウィルコクソンの順位和検定」、「クラスカル・ウォリスのH検定」があり、5)の方法には、「正規化順位法」がある。

1) スピアマンの順位相関係数

順位相関は、順序尺度データで与えられる2変数同士や、パネリスト同士の相関の程度を示すもので、スピアマンの順位相関係数^{12),22)}では、2つの特性それぞれに順位がつけられた時、その順位値をそのまま計量値とみなして、相関係数を計算する。

<入力：パネリスト同士の相関を求める場合>

表 2.2.1 データ例

パネリスト	試料 A	試料 B	試料 C	試料 D	試料 E	試料 F
1	1	2	4	3	6	5
2	2	3	4	1	5	6
d	-1	-1	0	2	1	-1
d ²	1	1	0	4	1	1

片側検定（+か-のいずれかの値しかとらないことがわかっている）か両側検定（+と-の値をとる可能性がある）かに注意する必要がある。

同順位がある場合は、平均順位をつける。

<同順位がない場合の r_s は、下記の式で求める>

$$r_s = 1 - \frac{6 \sum_i d_i^2}{p(p^2 - 1)} \quad (2.2.1)$$

r_s : スピアマンの順位相関係数

p: 試料の数

d: 試料 i に関して、2つの順位の差