

原 著

順位法を用いた視覚評価の信頼性について
—順序尺度の解析と正規化順位法による尺度構成法—

中前光弘

奈良県立医科大学附属病院中央放射線部

論文受付
1999年8月11日論文受理
2000年1月20日

Code No. 520

はじめに

昨今, computed radiography (CR)をはじめmagnetic resonance images (MRI), computed tomography (CT), digital subtraction angiography (DSA), digital radiography (DR) などあらゆる分野でデジタル画像が普及してきている。それに伴い, デジタル画像の特長を十分に引き出すために, どのような画像処理を行い臨床の場へ提供すればよいのか決める必要性がある。

臨床の場で画像の良し悪しを決めるのは, 臨床医である。人間の感覚器官である視覚によって, 最終判断が下されることから, 視覚評価法が用いられている。そのなかでも, 比較的多数の刺激(試料画像)を互いに関連させ, 判断が比較的容易にできるなどの簡便さか

ら, 試料を順位付ける方法(以下, 順位法)が頻繁に用いられている¹⁾。

たとえば, 観察者が同時に提示された試料画像5枚を, その良し悪しについて順位付けする場合を考える。この時観察者は, 「順位を付ける」という規則に従って, 5枚の試料画像の良さを測定したことになる。この規則のことを「尺度」という²⁾。1951年, Stevensが測定の定義のなかで, 尺度を「分類, 順序, 距離, 比例」の四つに分類している³⁻⁵⁾。これらの尺度の性質や許容される統計量は, 後者になればなるほど高度になり, それよりも低度の尺度の性質をすべて併せ持っていることになる。したがって, ある尺度を使って測定したデータに対しては, その尺度よりも低度の尺度を

Study of the Reliability of Visual Evaluation by the Ranking Method: Analysis of Ordinal Scale and Psychological Scaling Using the Normalized-rank Approach

MITSUHIRO NAKAMAE

Central Department of Radiology, Nara Medical University Hospital

Received Aug. 11, 1999; Revision accepted Jan. 20, 2000; Code No. 520

Summary

The method of visual evaluation is frequently used, as is the ranking method, because it is a relatively simple means of evaluation. However, in order to measure a ranking method according to the rule of "attaching ranking", results are given with an ordinal scale. The size relation of the scale is guaranteed, but the interval values. Therefore, the average value does not have any mathematical meaning. In order to compensate for these problems, we attempted to determine the right method of analysis for ordered data and examined the normalized-rank method, which can compute a more advanced distance scale, and its method of analysis. The concept of a scale was introduced in order to carry out analysis suitable for the purpose of visual evaluation. Consequently, the reliability of the ranking method was improved.

Key words: Visual evaluation by the ranking method, Ordinal scale, Distance scale, Normalized-rank approach, Test of significant

別刷資料請求先: 〒634-8522 奈良県橿原市四条町 840
奈良県立医科大学附属病院 中央放射線部 中前光弘 宛

使って測定したデータに対して適用できる統計量はすべて適用されることになる。

しかし、「順位を付ける」という規則に従って測定されたデータは順序尺度であり、その大小関係は保証されているものの、間隔の幅が未知で等しくない。そのため、順位の平均値は数学的に意義を持たないことになる⁶⁾。

ここでは、視覚を用いて試料を順位付けしただけで、評価を行ったことにはならない。結果を解析し、試料間の有意差を検定するなどの価値判断を付加することで「視覚評価」を行ったといえることができる。

そこで、本論文では、順序尺度である順位データの取り扱い方と正しい解析方法について検討した。そして、正規化順位法を用いて順序尺度からより高度な距離尺度を算出し、正しい解析方法について検討したので報告する。

1. 方法

試料画像として、頭部ファントムを用いて、管電圧70kV一定でmAs値を8, 12, 16, 20, 40(以下, A, B, C, D, E)と変化させて撮影した聴器(stenvers法)

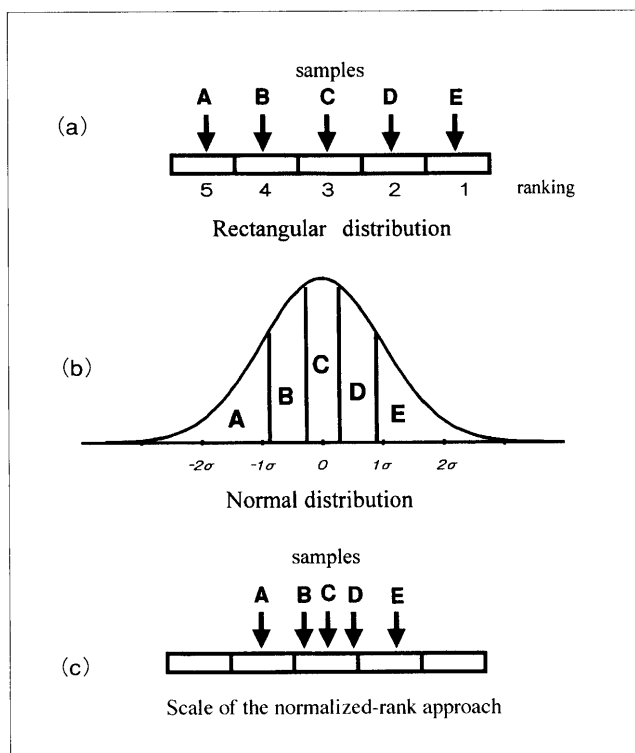


Fig. 1 Figures showing the process of psychological scaling using the normalized-rank approach.

(a) Figure shows the distribution of ordinal data of five samples.

(This can be regarded as a rectangular distribution of frequency 1.)

(b) The distribution when evaluating preference.

(The normal distribution is shown in approximation. Each area of A-E is equal.)

(c) Scale of the normalized-rank approach.

のCR画像5枚を用意した。

次に、5枚の試料画像を同時に提示し、放射線技師10名でその良し悪しについて順位付けをした。

1-1 順序尺度による解析

得られた順位データは順序尺度であり、前述したように平均順位を算出しても数学的に意味を持たないことになる。そこで、評価の指標として、各試料に対する順位の総和 R_i を求める。そして、 R_i から試料間の関係を解析するために順序尺度でみた多変量(関連多群)間の相関指標であるKendallの一致性の係数 W を求め、検定を行う⁷⁻⁹⁾。

1-2 正規化順位法による距離尺度への変換

順序尺度の場合、試料間に何らかの差があることはわかるが、どの試料とどの試料の間に有意差があるかが不明である。そこで、試料のどの区間に有意差があるのかを解析するためには、より高度な距離尺度に変換する必要がある。その手法として、正規化順位法^{6, 10, 11)}を用いているが、その概要を説明する。

順位データは、5個の間隔とおおのの間隔に度数1を持つ矩形分布と見なすことができる(Fig.1a)。この順位を数値と見なせば等距離であるが、普通はそうではない。対象が品質の良さなどの場合には、ほとんどが正規分布をなす母集団から出ている可能性が強く、一部に不規則性はあっても隣接した対象間の距離は、中心付近で小さく両端で大きいという「正規曲線の原理」に従っている⁶⁾。そこで、各対象が標準正規分布(平均0, 分散1の正規分布)の曲線下で、同じ大きさの面積を占めていると考えられる(Fig.1b)。N個の試料を抜き出した時に、第*i*番目の試料が占める位置(分布の中心から標準偏差の単位で測った距離)の期待値を計算し、これを*i*番目の試料にスコア(正規スコア)として与える(Fig.1c)。

1-3 距離尺度による解析

順序尺度を距離尺度(正規スコア)に変換し、各試料ごとの評点を推定する。また、試料と観察者を要因とする二元配置分散分析から試料間の有意差検定を行う。有意差があった場合は、どの試料とどの試料の間に有意差があるかを知るために、*t*分布を用いて有意差があるといえる差の最小値(least significant different: 以下, l.s.d.)を求め、有意差検定を行う^{12, 13)}。

2. 結果

2-1 順序尺度による解析

観察者10名による順位データおよび順位の総和 R_i をTable 1に示す。 R_i から判断すると $A < B < C < D < E$ の

Table 1 Ordinal data for five samples ranked by ten observers and total (Ri).

Observers	Samples					Total
	A	B	C	D	E	
a	5	3	4	2	1	15
b	5	4	2	3	1	15
c	5	4	3	1	2	15
d	4	5	1	3	2	15
e	4	3	5	2	1	15
f	5	4	1	3	2	15
g	5	4	3	2	1	15
h	2	5	3	4	1	15
i	5	4	3	2	1	15
j	5	4	3	1	2	15
Ri	45	40	28	23	14	150

Ri: Total of ordinal data

順になるが、その信頼性は低く順位を確定する根拠がない。

そこで、Kendallの一致性の係数 W を求める。算出のための補助表をTable 2に示すが、 $W=0.634$ となる。試料数 $k \leq 4$ の場合は、Friedmanの検定表を用いるが、今回は $k=5$ であるため χ^2 検定を行う。

$$\chi^2 = n(k-1)W$$

から、 $\chi^2=25.36$ となる。ただし、 n は観察者数、 k は試料数。自由度 $f=4$ 、危険率 $\alpha=0.05$ の χ^2 値は、 $\chi^2_{(4, 0.05)}=9.49$ で $\chi^2 > \chi^2_{(4, 0.05)}$ となるため、順位について一貫性があることになる。すなわち、試料間の順位には、何らかの差があることになる。

2-2 正規化順位法による距離尺度への変換

試料数 k が 5 の場合、正規スコアは、1 位 1.16, 2 位 0.50, 3 位 0, 4 位 -0.50, 5 位 -1.16 となる^{10, 11)}。順位データを正規スコアに変換したものをTable 3に示す。

2-3 距離尺度による解析

各試料における正規スコアの平均が、正規化順位尺度つまり距離尺度となる(Table 3の最下段)。

次に、二元配置の分散分析表をTable 4に示す。分散 $F_A=14.88$ となり、自由度 $f_A=4$, $f_e=36$ 、危険率 $\alpha=0.05$ の F 値、 $F_{(4, 36, 0.05)}=2.63$ より大きいため、試料間に有意差がある。

それでは、どの区間に有意差があるか、危険率 $\alpha=0.05$ の $l.s.d.$ を求める(Table 4)。 $l.s.d.(5\%)=0.52$ であり、これ以上離れた区間が有意差ありということになる。

距離尺度を数直線状に表し、同時に試料間の有意差を示したものをFig.2に示す。

Table 2 Auxiliary table and expression for computing the coefficient of concordance of Kendall (W).

Observers	Samples					Total
	A	B	C	D	E	
a	5	3	4	2	1	15
b	5	4	2	3	1	15
c	5	4	3	1	2	15
d	4	5	1	3	2	15
e	4	3	5	2	1	15
f	5	4	1	3	2	15
g	5	4	3	2	1	15
h	2	5	3	4	1	15
i	5	4	3	2	1	15
j	5	4	3	1	2	15
Ri	45	40	28	23	14	150
Ri-R	15	10	-2	-7	-16	0
(Ri-R) ²	225	100	4	49	256	634

$$\Sigma Ri = 150$$

$$R = n(k+1)/2 = 10 \times (5+1)/2 = 60/2 = 30$$

$$S = \Sigma (Ri - R)^2 = 634$$

$$W = 12S / [n^2 (k^3 - k)] = 12 \times 634 / [10^2 \times (5^3 - 5)] = 0.634$$

n: observers, k: samples

S: Sum of squares

W: Kendall's coefficient of concordance

Table 3 Table used to convert ordinal data into scores for ordinal data.

(The average value of the scores for ordinal data of each sample are given.)

Observers	Samples				
	A	B	C	D	E
a	-1.16	0.00	-0.50	0.50	1.16
b	-1.16	-0.50	0.50	0.00	1.16
c	-1.16	-0.50	0.00	1.16	0.50
d	-0.50	-1.16	1.16	0.00	0.50
e	-0.50	0.00	-1.16	0.50	1.16
f	-1.16	-0.50	1.16	0.00	0.50
g	-1.16	-0.50	0.00	0.50	1.16
h	0.50	-1.16	0.00	-0.50	1.16
i	-1.16	-0.50	0.00	0.50	1.16
j	-1.16	-0.50	0.00	1.16	0.50
Total	-8.62	-5.32	1.16	3.82	8.96
Mean	-0.86	-0.53	0.12	0.38	0.90

順位は $E > D > C > B > A$ となるが、 $E-D$, $D-C$, $B-A$ 間には有意差が認められず、順位の入替わる可能性は否定できない。

3. 考 察

順位法は、観察試料を同時に提示するため、観察者にとっては評価しやすい手法であるが、データを解析する場合には、前述したように順序尺度として取り扱う必要がある。

視覚を用いて「測定」するのは簡単である。しかし、

Table 4 Table of analysis of variance and expression of least significant difference (level of significance, 5%).

Factorial variation	Degrees of freedom f	Sum of squares S	Variance V	Ratio of variance F
Total (T)	fr=49	ST=31.91		
Factor (A)	fa=4	SA=19.88	VA=4.97	FA=VA / Ve=14.88
Factor (B)	fb=9	SB=0	VB=0	
Error (e)	fe=36	Se=12.03	Ve=0.33	

$$F_0 = F(f_A, f_e; \alpha) = F(4, 36; 0.05) = 2.63 \quad \therefore F_A > F_0$$

$$sd = \sqrt{(2Ve / n)} = 0.257$$

n: sample, sd: standard deviation

$$l.s.d. (5\%) = t(f_e; \alpha) \times sd = t(36; 0.05) \times sd = 0.520$$

l.s.d.: least significant difference

$$t(36; 0.05) = 2.028$$

α : level of significance

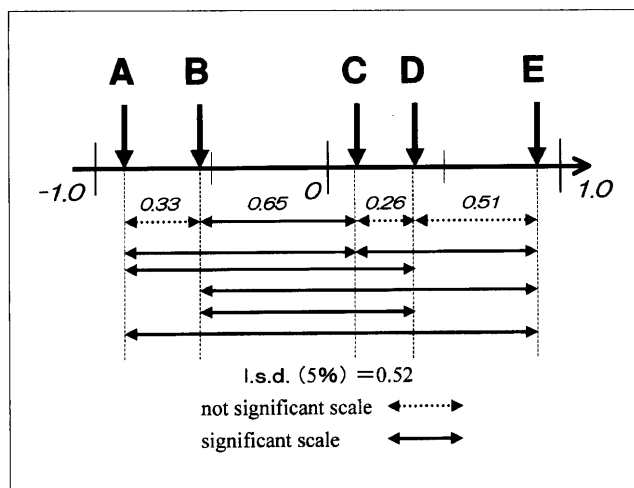


Fig. 2 Distance scales and 95% confidence interval. Ten observers ranked five sample images. The ordinal data changed distance scales using the normalized-rank approach.

測定結果に客観性を持たせ、どのように価値判断を含めるかが、「評価」の難しい点であり、実験の目的にあった「評価」を行うには、どのような「尺度」を必要としているかを考え、その「尺度」に合わせた正しい解析手順を実施する必要がある。その流れ図をFig.3に示す。

何らかの物理量と画像の順位関係を調べることが目的であれば、試料間の順位について検定を行えば問題はない。しかし、X線画像の評価に欠かせない試料間の有意差を求めることはできない。そのため、われわれは順序尺度を距離尺度に変換できる正規化順位法を積極的に用いてきた^{12, 14-16)}。そして、距離尺度からl.s.d.を求め有意差検定を行っている。

今回の実験では、試料として頭部ファントムの聴器CR画像を用いている。線量の変化は、画像のノイズ成分の増減を意味している。線量(mAs値)の変化と画質の関係を調べる場合、順序尺度についてKendallの一致性の係数を求め、検定の結果が有意であれば

Spearmanの順位相関係数 r_s を求めればよい。 $r_s=1.00$ となり、8~40mAsの間では、線量を増やせば画質も良くなるのが分かる。しかし、患者の被曝線量を軽減できる最適な線量を求める場合、それぞれの試料間の有意差を求めなければ結果を評価できない。そこで、順位尺度を距離尺度に変換し、どの試料間に有意差があるかを求める。8(A), 12(B)mAsは、16(C), 20(D), 40(E)mAsに比べて有意差があり、明らかに劣っていたが、16(C), 20(D), 40(E)mAsでは、C-DとD-E間に有意差がなかった。患者の被曝を考慮すると16mAs程度で撮影しても画質には大差がなく、最適線量を導き出すことができる。

ただし、正規化順位法がすべてのケースに当てはまるものではなく、二、三の問題点を解決してはじめて有意義な手法となることを忘れてはいけない。たとえば、順位法自体が持つ問題として、同順位の問題が考えられる。つまり、順位が同じ場合にどのような処理をするかについて、何らかの仮定を設ける必要がある。今回の実験では、「同順位を認めない」という測定方法を採用している。そのため、試料間の刺激差が非常に小さい場合には、この手法では検出力が低く、不向きである。このような場合は、相対的な評価で小さな差を検出できる一対比較法¹⁷⁻¹⁹⁾など別の手法を用いるべきである²⁰⁾。

また、測定の要素についても考えなければならない。これには、観察者(評価者)、試料(評価対象)、評価基準がある。正規化順位法の場合、観察者数と試料数については実験で統制しているが、評価基準すなわち観察者の能力がある程度安定していないと「正規分布」に近似する前提が成り立たないことになる。したがって、嗜好形の評定実験であっても分析形に準拠する必要がある。観察者の選定および実験前学習が結果を左右することを認識しておくことが重要

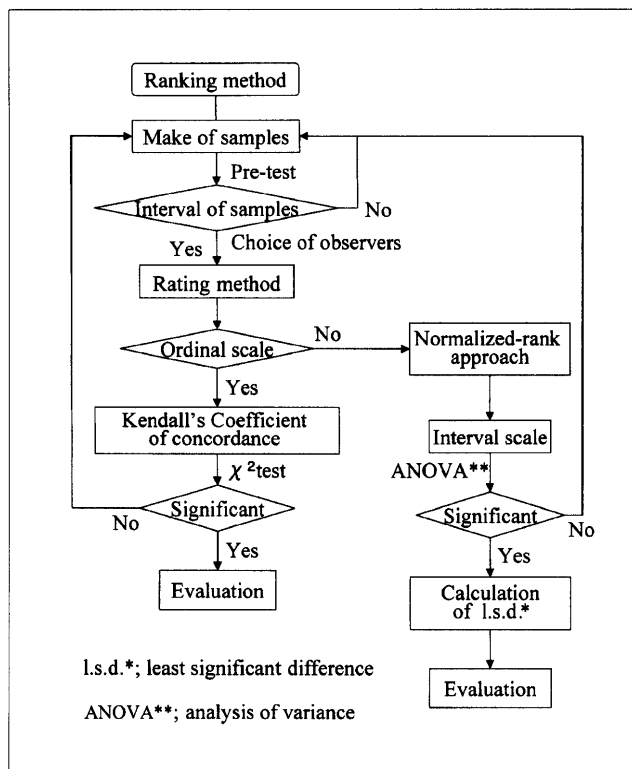


Fig. 3 Flowchart for the ranking method.

である^{20, 21)}。

そして、試料の問題も忘れられない。あまりに数が少なすぎても「正規分布」が成り立たないし、多すぎると上位や下位の順位付けができて中位の順位付けは非常に困難な作業となる。そこで、一般的には5～10程度の試料数とするのが望ましい。

それから、刺激の大きさも重要な因子である。試料間の刺激差があまりに大きすぎると感度が低下する場合がある。Fig.2の結果を例にとって説明すると、A、BはともにCから有意差があり、A、Bの順位は両者の関係から算出されている。ここで、Bを除いたA_B、C_B、D_B、E_BとAを除いたB_A、C_A、D_A、E_Aの順位データから距離尺度値を算出し、Fig.4の数直線上に示す。C_B(C_A)からみると、A_BもB_Aも有意差があり、同じくらいの距離である。ここでいえることは、正規化順位法とは、順位データを基に距離尺度を算出する方法で

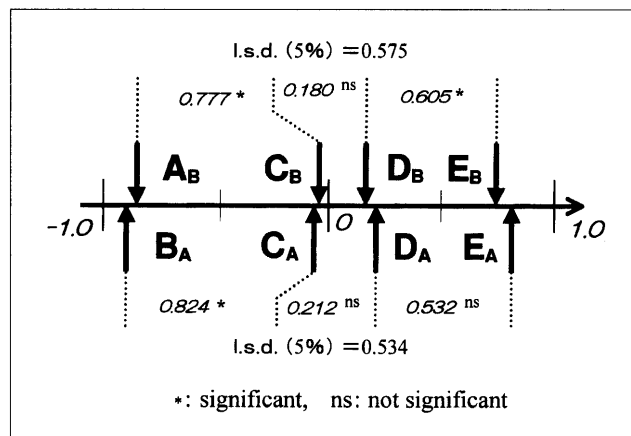


Fig. 4 Distance scales and 95% confidence interval. Order data when removing sample A or B, change distance scales using the normalized-rank approach.

あるために、隣接する試料から有意に離れた点については、有意に離れていることは分かっても正確な距離を把握できていないことになる。特に、端試料の場合は、隣接する試料が1点しかないため、注意が必要である。したがって、予備実験で試料間の刺激の大きさを把握してから、本実験を実施する必要がある。

4. まとめ

順位法で求められた順序尺度の正しい解析方法を使用し、結果について評価を行った。また、順序尺度を正規化順位法によって距離尺度に変換した。その距離尺度を使って、試料間の優劣はもちろんのこと、有意な差がある距離(l.s.d.)を求めて、どの試料とどの試料の間に有意差があるかの検定を行った。

視覚評価の目的にあった解析をするために、尺度の概念を導入することで、順序データを扱う順位法の信頼性が向上した。

謝 辞

稿を終えるにあたり、ご協力ならびにご助言をいただきました当院中央放射線部 越智 保技師長ならびに大賀泰文副技師長に御礼申し上げます。また、観察者としてご協力いただいた当院技師諸兄に感謝いたします。

参考文献

- 1) 中前光弘, 森岡雅幸, 和氣利明, 他: 視覚評価の現状と問題点について. 日放技学会近畿部会誌, 4(2), 13-17, (1999).
- 2) 大平 平: 評価と数量化のはなし. pp.23-35, 日科技連出版社, 東京, (1983).
- 3) 佐藤 信: 官能検査入門. pp.129-135, 日科技連出版社, 東京, (1978).
- 4) 日科技連官能検査委員会 編: 新版 官能検査ハンドブック. pp.243-246, 日科技連出版社, 東京, (1973).
- 5) 市原清志: バイオサイエンスの統計学. pp.8-11, 南江堂, 東京, (1985).
- 6) Guilford JP: 秋山義治 監訳: 精神測定法. 培風館, 東京, (1959).
- 7) 佐藤 信: 官能検査入門. pp.63-73, 日科技連出版社, 東京, (1978).
- 8) 市原清志: バイオサイエンスの統計学. pp.244-245, 南江堂, 東京, (1985).
- 9) 増山栄太郎, 小林茂雄: センソリー・エバリュエーション—官能検査へのいざない—. pp.136-141, 垣内出版株式会社, 東京, (1989).
- 10) 佐藤 信: 官能検査入門. pp.74-76, 日科技連出版社, 東京, (1978).
- 11) Harter HL: Expected values of normal order statistics. *Biometrika*, 48, 151-165, (1961).
- 12) 田畑洋二, 大賀泰文, 角田充弘, 他: X線写真の主観的評価法—主観的尺度値の信頼性—. 日放技学誌, 49(1), 12-17, (1993).
- 13) 奥野忠一, 芳賀敏郎: 実験計画法. pp.9-12, 56-60, 培風館, 東京, (1969).
- 14) 中前光弘, 森岡雅幸, 和氣利明, 他: CR撮影における適正条件の検討—聴器撮影の線量最適化の実践—. 日放技師会誌, 44(12), 1718-1723, (1997).
- 15) 中前光弘, 池口俊孝, 澤井一郎, 他: 固形薬品処理システムによる自動現像機TCX-201の臨床評価. コニカX-レイ写真研究, No.245, Vol.49(4), 120-127, (1998).
- 16) 川口真一, 下村 学, 細木秀彦, 他: パノラマX線画像の視覚評価に関する基礎的検討. 四国歯誌, 11(2), 253-262, (1999).
- 17) 大賀泰文, 田畑洋二, 辻本武士, 他: X線写真の主観的評価法—Thurstoneの一対比較法による尺度化の試み—. 日放技学誌, 45(7), 831-839, (1989).
- 18) 中前光弘, 田畑洋二, 大賀泰文, 他: Schefféの一対比較法による主観的評価法. 日放技学誌, 52(11), 1561-1565, (1996).
- 19) 大塚昭義, 砂屋敷忠, 小寺吉衛 編, 吉田 彰 著: 実験画像評価—初学者のための実験入門書—. pp.81-96, 日本アクセル・シュプリング出版, 東京, (1994).
- 20) 井上裕光: 私信, (1998).
- 21) 中前光弘, 森岡雅幸, 河合寿夫, 他: 視覚評価の信頼性について—分析形および嗜好形評定実験の特性と問題点—. 日放技学誌, 53(10), 1525-1529, (1997).

図表の説明

- Fig. 1 正規化順位法による尺度化の課程.
 (a) 5 個の順位データの分布を示す図(度数 1 の矩形分布とみなすことができる).
 (b) 良し悪しを評価した時の分布図(近似的に正規分布を示す. A~Eの各面積は等しい).
 (c) 正規化順位尺度値の分布.
- Fig. 2 観察者10名で 5 枚の試料画像を順位付けし, 正規化順位法で距離尺度に変換した値. 95%信頼区間(危険率 5%での有意差)も示している.
- Fig. 3 順位法の流れ図.
- Fig. 4 試料AおよびBを除き, 4 枚の試料画像の順位データから正規化順位法で変換した距離尺度値. 95%信頼区間(危険率 5%での有意差)も示している.
- Table 1 観察者10名が 5 枚の試料画像に順位付けをした順位データとその総計 R_i .
- Table 2 Kendallの一致性の係数 W を算出する補助表と計算式.
- Table 3 順位データを正規スコアに変換した表(各試料の正規スコアの平均値が, 結果になる).
- Table 4 正規スコアから求めた分散分析表と $t.s.d.(5\%)$ の算出式.