医学教育を考える2
試験の識別指数

川崎医科大学情報科学

近　藤　芳　朗

(平成20年12月18日受理)

The Role of the Discrimination Index on Tests

Yoshiro KONDO

Department of Information Sciences, Kawasaki Medical School,
577 Matsushima, Kurashiki, Okayama, 701-0192, Japan
(Received on December 18, 2008)

概　　要

識別指数ϕはふつう個々の小問題に対して算出されるのであるが，本論文では全部の小問題についての平均識別指数$\bar{\phi}$を試験全体の識別指数とみなして試験の識別能力について論じた。その結果，平均識別指数$\bar{\phi}$は，試験の平均正答率をμ，標準偏差をσとすると$\bar{\phi}=1.3\sigma/\sqrt{\mu(1-\mu)}$という関係のあることがわかった。また，全受験者数を$2N$とすると$\chi^2=N\phi^2$が成り立つので，$N=60$人の場合，有意水準5％で，試験によって上位と下位が識別されるためには$\phi>0.253$であればよく，このことから標準偏差は100点満点で10点以上あればよいことがわかった。
キーワード：識別指数，平均点，標準偏差，χ^2分布

Abstract

The discrimination index ϕ is usually evaluated on each problem, but the mean discrimination index $\bar{\phi}$, which is concerned with all the problems on the test, is discussed in the present study. As a result, it was found that the mean discrimination index $\bar{\phi}$ was equal to $1.3\sigma/\sqrt{\mu(1-\mu)}$, where μ denotes the average rate of correct answers on the test and σ denotes the standard deviation. From this, it was concluded that the condition of $\phi>0.253$ and $\sigma>0.1$ is necessary to discriminate a higher level group from a lower level group by the test on the significant level of 5%.
Key words: discrimination index, mean value, standard deviation, chi-squared distribution

1．はじめに

試験は「競争試験」と「資格試験」とに大別できる。「競争試験」は「選抜試験」とも呼ばれ合格者定員が決まっているもので，受験者の質にかかわらず成績上位の一定数の者が選抜される競争の試験である。入学試験がその典型である。これに対して「資格試験」は合格者定員

Correspondence to Yoshiro Kondo

などなく，一定のレベルに達していれば，つまり合格点が決められているので，何人でも合格できる。したがって，受験者全員合格ということもあり得るし，また，全員不合格ということもある。各種の国家試験や大学の進級試験などは「資格試験」である。

競争試験の場合，試験毎の難易度の変動は問題にならないが，資格試験の場合は問題となる。この場合，試験の難易度に応じて合格基準を変える必要もでてくる。試験問題の良否は成績上位者と下位者を識別できるか否かにある。上位者と下位者を識別できる問題は競争試験では特に良い問題となる。個々の問題についての上位者と下位者を分離する分離能力の指標を識別指数という。資格試験については適性な合格基準が求められるが，これは試験問題の難易度にも左右される。この論文では「識別指数」と「合格基準」について論ずる。

2．識別指数（discrimination index）

医歯薬系共用試験CBT（Computer Based Testing）で用いられている識別指数ϕは次式で定義されている。

$$\phi=\frac{(A\times D-B\times C)}{\sqrt{(A+B)(C+D)(A+C)(B+D)}} \tag{1}$$

ここに，

A＝総得点が上位1/4にあるグループの当該問題正解者数
B＝総得点が上位1/4にあるグループの当該問題誤答者数
C＝総得点が下位1/4にあるグループの当該問題正解者数
D＝総得点が下位1/4にあるグループの当該問題誤答者数

である。この識別指数は問題ごとに算出される。上位グループ，下位グループの正解者数・誤答者数の値によってϕの取り得る値の範囲は

$$-1\leqq\phi\leqq 1 \tag{2}$$

となる。たとえば，

上位1/4が全員正解者で下位1/4が全員誤答者である場合：$\phi=1$
上位1/4が全員誤答者で下位1/4が全員正解者である場合：$\phi=-1$
上位と下位で正解者の数が同じである場合　：$\phi=0$

である。$\phi=1$では上位と下位が完全に識別されているのに対して，$\phi=0$では上位と下位が全く識別できていない。この識別指数ϕは統計学的には自由度１のχ^2と

$$\chi^2=N\phi^2 \tag{3}$$

の関係がある。ただし，$N=A+B+C+D$である。また，このϕは，正答，誤答に応じて１，０の２値的なデータを与えるときの上位・下位グループについてのピアソンの相関係数にも一致している。また，教育学ではϕ（ファイ）係数と呼ばれている。

ϕがどのような値をとれば上位と下位のグループが有意に差があるといえるのだろうか。それは式(3)からもわかるように有意水準αで上位と下位のグループが識別されるためには

$$\phi > \frac{\chi^2_\alpha}{N} = \phi_0 \tag{4}$$

であればよい。N=60人の場合（全受験者は120人，これは本学のほぼ各学年の学生数に等しい）となる。この表から，有意水準5％で上位1/4と下位1/4が識別できるためには120人の受験者に対してϕは0.253より大きければよいことがわかる。

表1　有意水準αと識別指数ϕ_0

α	0.1	0.05	0.025	0.01
χ^2_α	2.71	3.84	5.02	6.63
ϕ_0	0.213	0.253	0.289	0.332

識別指数は式(4)からわかるように受験者数に依存する。受験者数2Nが多ければ多いほど識別指数は小さくても良い。このことは上位者と下位者の成績が区別がつけられないような集団と比べての話しであってこのような集団では得点の分散は1/Nに比例して小さくなるのであるが，実際の集団では100人の集団でも1000人の集団でもまた，センター試験のように数十万人の集団でも総得点の分布は変わらず分散はほぼ同じである[1)]。しかし，この集団が数人とか数十人の場合は得点分布は不安定で分散も大きい。したがって，全受験者が100人程度での統計学的議論は識別指数に限りより多くの受験者の場合の標準となる。このような議論を踏まえて識別指数については受験者の数に無関係に

$\phi > 0.5$ … 稀（良）

$\phi \geqq 0.25$… 適当

$\phi \leqq 0.15$… 質的に疑問

とされている[2)]。

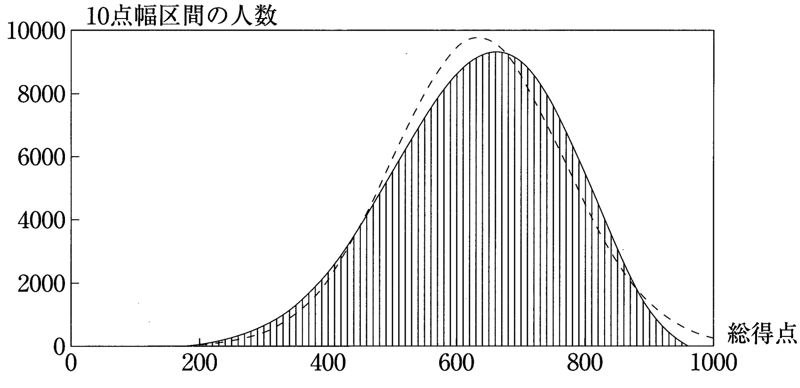

図1　昭和51年度の共通一次試験の総得点の分布[1)]

実線はμ＝636.07, σ＝134.28のベータ分布を表し、破線は同じμ, σの正規分布を表す。

3．平均識別指数 $\bar{\phi}$ と平均点 μ および標準偏差 σ の関係

識別指数 ϕ は個々の小問に対して算出されるべきものであるが，ここでは試験全体が上位者と下位者を識別するのに妥当か否かについて考える。試験はm個の小問からなり，それぞれの小問は正答か誤答かであり部分点はないものとする。この試験において上位1/4の者のm個の小問すべてについて延べの正答者数をaとし，１問あたりの平均の正答者数をAとすると

$$A=a/m \tag{5a}$$

である。同様にして，上位1/4の者の m 個の小問すべてについての延べの誤答者を b とし，１問あたりの平均誤答数を B とすると

$$B=b/m \tag{5b}$$

であり，下位1/4の者の m 個の小問すべてについての延べの正答者数を c，誤答者数を d とし，１問あたりの平均の正答者数，誤答者数をそれぞれ C および D とすると

$$C=c/m,\ D=d/m \tag{5c}$$

となる。これらの A, B, C, D に対して式(1)で算出される値を平均の識別指数と呼び $\bar{\phi}$ で表す。平均識別指数 $\bar{\phi}$ は以上の定義からわかるように個々の小問の識別指数
$\phi_i(i=1,2,\cdots,m)$ の平均ではない。すなわち

$$\bar{\phi} \neq \frac{1}{m}\sum_{i=1}^{m}\phi i \tag{6}$$

である。$\bar{\phi}$ についても表１はそのまま利用できる。

この試験において，成績上位1/4の者の平均正答率を μ_1，成績下位1/4の者の平均の正答率を μ_2 とすると平均識別指数 $\bar{\phi}$ は，

$$A+B=C+D=N/2 \tag{7}$$

に注意すると式(4)から

$$\begin{aligned}\bar{\phi} &= \frac{AD-BC}{\frac{N}{2}\sqrt{(A+C)(B+D)}}\\ &= \frac{\mu_1-\mu_2}{\sqrt{(\mu_1-\mu_2)(2-\mu_1-\mu_2)}}\end{aligned} \tag{8}$$

が得られる。いま，受験者全員の正答率分布について，平均値(平均正答率) を μ，分散を σ^2 とし，この μ，σ^2 と $\bar{\phi}$ の関係を求める。

a．受験者の正答率分布が正規分布に従う場合

この場合，

$$\left.\begin{aligned}\mu_1&=\mu+\frac{4\sigma}{\sqrt{2\pi}}e^{-\frac{1}{2}(0.6745)^2}\\ \mu_2&=\mu-\frac{4\sigma}{\sqrt{2\pi}}e^{-\frac{1}{2}(0.6745)^2}\end{aligned}\right\}\tag{9}$$

が得られる。式(8)，(9)から

$$\begin{aligned}\bar{\phi}&=\frac{4\sigma}{\sqrt{2\pi}}e^{-\frac{1}{2}(0.6745)^2}/\sqrt{\mu(1-\mu)}\\ &=\frac{1.2711\sigma}{\sqrt{\mu(1-\mu)}}\end{aligned}\tag{10}$$

が得られる。

b．受験者の正答率分布が三角分布に従う場合

正答率分布が右図のように（$\mu-\Delta$）と（$\mu+\Delta$）の間に対称的に分布している場合，

$$\left.\begin{aligned}\mu_1&=\mu+\left(1-\frac{\sqrt{2}}{3}\right)\Delta\\ \mu_2&=\mu-\left(1-\frac{\sqrt{2}}{3}\right)\Delta\end{aligned}\right\}\tag{11}$$

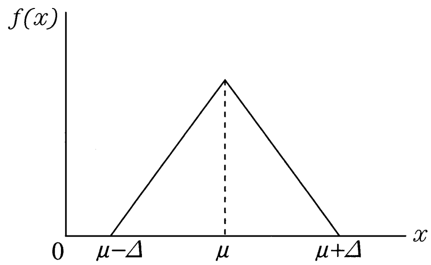

図2　三角分布

が得られるので，式(8)，(11)から

$$\bar{\phi}=\frac{1.2947\sigma}{\sqrt{\mu(1-\mu)}}\tag{12}$$

が得られる。　ここで，σ^2は分散で

$$\Delta=\sqrt{6}\,\sigma\tag{13}$$

の関係があることを用いた。

以上のことから受験者の正答率分布が正規分布でなくても上に凸の釣鐘状の分布をしていれば近似的に

$$\bar{\phi}=\frac{1.3\sigma}{\sqrt{\mu(1-\mu)}}\tag{14}$$

が成り立つといってよい。この式(7)からわかることは平均正答率μが同じならば$\bar{\phi}$は標準偏差σに比例する。σはバラツキの度合いを表すのだからこの結果は当然である。次にσが同じならばμが0かあるいは1に近いとき，$\bar{\phi}$は非常に大きくなる。しかし，ϕが0か1に近いとき，一般にはσが小さくなる可能性が高いので結果として$\bar{\phi}$は必ずしも大きくなるとは限らない。したがって，σを一定に保ちながらμを大きく，または小さくすることができれば該当の試験は識別能力の高い良い試験と言えるのである。

σが一定であれば識別指数が最も小さくなるのは$\mu=1/2$の場合で，このとき，

$$\bar{\phi} = 2.6\sigma$$

となる。$\sigma = 0.1$ であれば $\bar{\phi} = 0.26$ となり，有意水準５％で上位と下位のグループが識別できることになる。これは100点満点の試験に換算すると，上位と下位のグループを有意水準５％で識別するには標準偏差 σ は10点はなければならないということを意味する。

結論として，選抜試験では標準偏差は大きくなければならない。資格試験では平均値が高く，かつ標準偏差も大きい方がよい。資格試験で全員が合格するような集団に対しても標準偏差が大きくすなわち識別指数が大きい分離能力のある問題を出題すべきで，そのような試験で欠点をとった学生は不運ではなく，正しく・実力で欠点を取ったことになるからである。

４．資格試験ついての合格基準の設定

資格試験の合格基準が100点満点の60点であるとき(多くの場合そうだが) かつ適切な試験問題が出題された場合当然合格基準を規則通り60点として良いことは明らかだが，出題された問題の難易度が高く，多くの不合格者が生ずる場合の合格基準を60点とすると受験生には不利となる。本来，適切な水準の問題が出題されている場合。標準レベルの受験者ならば平均点が例えば75点とする。いま，当該の試験での平均点が65点であるとすると，これを標準の75点に引き上げるためには受験者全員の得点を（75/65）倍すればよい。こうしておいて60点を合格基準にすれば文句はでまい。この引き上げられた合格基準の60点は素点では $65 \times (60/75) = 52$ 点となる。実際は，素点を引き上げたりしないで合格基準を52点に下げるだけで処理は済むが，規則上60%以上が合格となっていれば実際（75/65)倍する。一般に平均点を μ とすると，合格点は $\mu \times (60/75) = 0.8\mu$ となる。つまり，平均点の80% が合格点となるのである。

ドイツの医師国家試験の合格点の設定は上述の方法に似ている。合格の判定は次のように記されている。「出題数60%以上が正解であれば合格，または最短年限で受験した受験生の正解数の平均から22%以内であれば合格となる。」となっている。つまり，最短年限の受験者の平均点の78%が合格基準なのである。これを，川崎医科大学の総合試験の合格基準に適用すると「60点以上であれば進級，または，一度も留年しないで進級した学生の平均点の78%以上であれば進級」となる。問題は78%が適切か，80%が適切かあるいは75%が適切かということである。逆に言えば問題は標準レベルの問題の標準学生の平均点をいくらと見るかということに帰着される。各学年の期末試験の平均点が70点～73点ぐらいであることを考えると少し高めの75点を標準の平均点とみて，すなわち合格点は平均点の78%ないしは80%が妥当ではないかと考えている。

次に，大学入試センター試験について，理科などの選択科目で難易度が極端に異なる場合，受験生には不公平感が残る。そこで，センター試験では次のような得点調節を行うことになっている。「物理，化学，生物，地学の平均点で一番高いものと一番低いものの差が20点を超え，これが試験問題の難易差によると認められる場合に得点調整を行うとしている。得点調整にあたっては最高平均点科目と最低平均点科目の平均点差が15点となるよう調整する。その際，平

均点がその間にある科目についても素点の平均点差の比率に応じて調整する」，となっている。この得点調整の方法は，各科目の受験者のレベルは同等と考えて，同じ分位点（パーセント点）の受験者は同じ能力と考えて調整するのである。つまり，最高平均点科目の累積分布を目標分布とし，調整すべき科目の累積分布を目標分布の方向へ移動させる方法，すなわち，分位点差縮小法を用いて行うのである（図３参照）。ここで重要なことは最高平均点科目の得点は不動とし，他の科目を引き上げるのである。こうすることによって引き下げられるよりは受験者の心理はやわらぐ。センター試験では20点差以上を15点にまでしか縮めないのであるが，これも10点差まで縮めるとか５点差まで縮めるとかの議論の余地はある[3]。

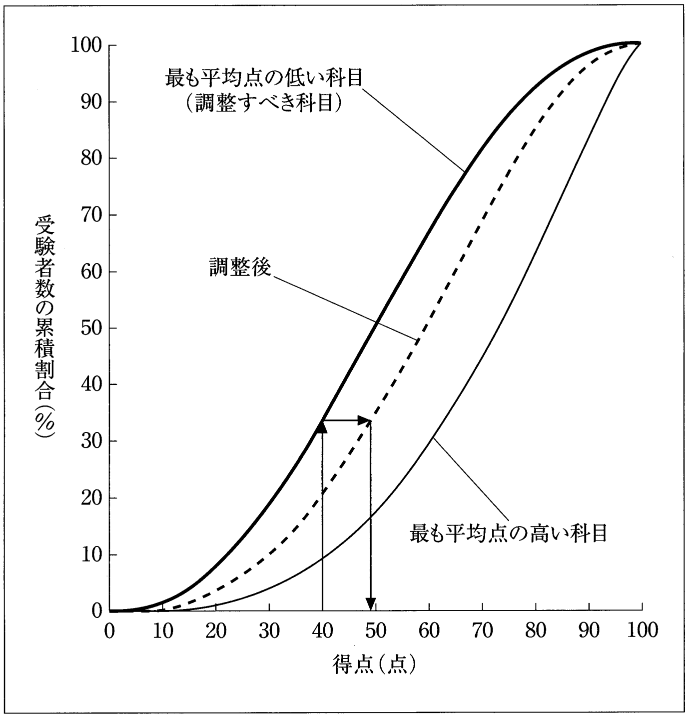

図３　得点の累積分布図
出典：「大学入試センター試験における得点調整について」
平成９年11月20日　大学入試センター

５．おわりに

識別指数，合格基準について述べたが，医学教育者にとって重要なことは，担当するそれぞれの科目でCBTのコアカリキュラムを含み，教えるべきミニマム・スタンダードを設定して講義を行い，また，大変な困難なことではあるが下位の学生のレベルアップを計ると共に上位の学生の意欲を殺ぐわないようにできれば最高である。そして適切な試験を実施することにより適切に評価し，学生は学生で反省し，教師は教師で反省すればより良い教育が達成できる。上

下分離能力の良い試験とは平均点が80点程度で（CBTでは各問の正答率80%を目標としている）で標準偏差が10点以上であるものをいう。

文　献

1）森口繁一：第６章試験の点数.『確率表現関数』. 東京大学出版会. 1995, pp81

2）「臨床実習開始前の学生評価のための共用試験システム」専門委員会：平成13年度 臨床実習開始前の学生評価のための共用試験システム CBT問題作成マニュアル. 2001, pp 6

3）阿部利則，近藤芳朗，山口恒夫：大学センター試験における得点調整方法— 合否入替り率による評価の試み—. 川崎医療短期大学紀要27：65－70, 2007