

# 総合卒業試験の識別性能向上と技術革新のための データ解析システムの開発

平野 光昭

入試改革の成果として、6年間でストレートに医師となる者の数が増えるとともに、人間性豊かな将来性のある医師を誕生させることができれば、改革は成功と言えようが、世間では国試(医師国家試験)の合格者を大学教育の1つの評価基準としているので、これを高めることにも無関心ではられない。そこで、国試と密接な関係のある総合卒試(総合卒業試験)の問の識別性能を点検するため、かつ国試の可否を予測するためのデータ解析システムを開発し、3年間にわたり改良を加えてきた。本論文では、どのような解析方法並びにコンピュータ・ソフトが開発されたかを紹介し、これらの方法を適用して、94年卒～98年卒の者を対象に行われた総合卒試のデータを解析した結果を報告する。もし、同様なテストが行われている全国の大学の協力が得られれば、識別性能の高い多数の問をコンピュータに蓄積することによって、すでに米国で行われているような、学生がコンピュータの前に座ってテストを受けることも夢ではなくなる。

キーワード：総合卒業試験，医師国家試験，識別性能，可否予測，試験技術革新

## 1 はじめに - 入試改革の成果を計る物差し

本学は、94年の推薦選抜の導入、その追跡調査に基づく99年の推薦選抜募集人員の増員など、開校(80年)以来常に入学者選抜方法の改善に努めているが、国際化、情報化、18歳人口の減少などの社会の変化や、他大学の改革の影響も絶えず受けて、入学してくる学生の質は、年年変化している。97年から分離分割方式を採用したが、前期日程入学者と後期日程入学者の入学後の成績を比較して、「平均値の差が有意である」と分かり、前期の方が高かったとしても、「本学の前期は募集人員が少ない」、「前期で合格しなかった者が後期を受けている」など、逆に後期の方が高かったとしても、「関東地区の医学科全体では前期の募集人員の方が圧倒的に多い」、「後期受験者の多くは前期で本学より難関な大学を受けている」など、主として選抜方法とは別な原因によって差が生じている可能性が高いため、これだけで両選抜方法に優劣を付けるわけにはいかない。また、国立大学の受験機会を1回にすれば、一般に県内高校出身者が増えるが、学内成績あるいは国試(医師国家試験)の成績を比較して、「県内高校出身者は県外高校出身者より平均して良い」という結果が出たとしても、「入学者100人の中に県内出身者が多い程その学年が全体として良い」ということには直接結びつかない<sup>(2)(3)(4)</sup>。

改革が「成功」と言えるか否かをみるには、このようなカテゴリー間の比較(図1のD)ではなく、改革前と改革後(図1のA, B, C), すなわち、入学年度間を比較する長い年月の追跡調査が必要である。しかし、医学科では大部分の授業が学年単位で行われているので、学内成績は集団内(同一学年)での相対評価となる傾向が強

い。そこで、入試改革の成果を計る(当然6年間の教育の成果も計られる)客観的な物差しとして、国試の大学としての成績、すなわち、合格率あるいは全国の大学の中でのその順位が注目される。もちろん、医学教育の目標は国試の合格率を高めることだけにあるわけではな

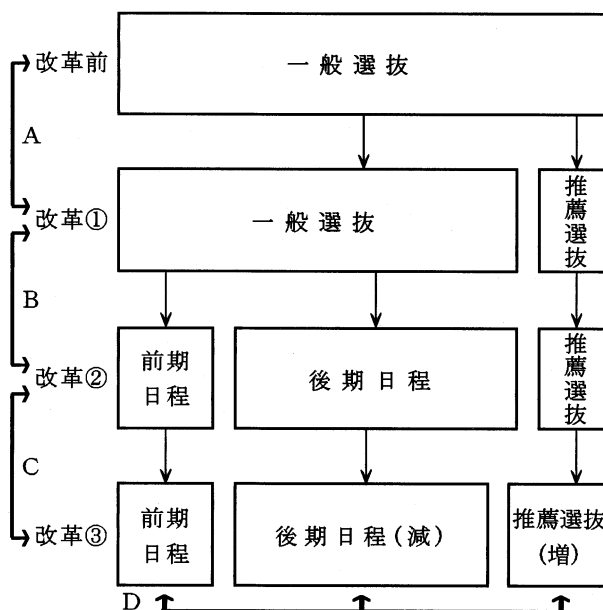


図1 入試改革と入学後の成績の比較

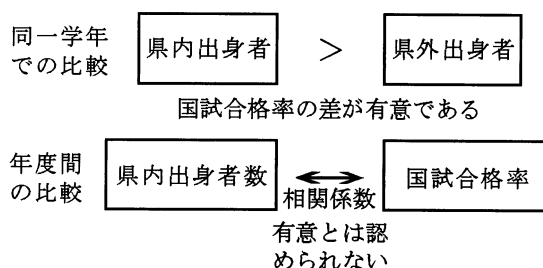


図2 集団内の比較と集団間の比較

く、各大学にそれぞれの教育目標があるわけであるから、本学の順位が入試改革の前後でどう変わったかを目安とするだけである。ところが、世間でもこの順位を医学教育の評価基準の1つとしているので、これを高める対策を講じている大学も多く、一般に卒業判定を厳しくすることによって、これを高めることができる。故に、国試を物差しとするのなら、6年間でストレートに卒業し、その年に国試にも合格した者の入学者に対する比率(ストレート合格率)を用いるべきである。

ところで、入学者の質を高めること及び教育の質を向上させることによって、ストレート合格率を高め、幅広い教養を身に付けた人間性豊かな良き医師、将来性のある医学研究者を誕生させることができれば、入試改革及び教育改革の目的は達せられたことになるが、医学科を卒業することを前提にして行われる国試も、「一発勝負」の試験であるから、「ストレート卒業生数を多くすれば、ストレート合格率は高まる」と一般には考えられている。大学審議会の「21世紀の大学と改革の方策について」と題する中間報告でも、卒業認定を厳しくすることがうたわれているが、ストレート合格率を入試改革の成果を計る物差しとするのなら、進級、卒業の判定基準をはっきりさせ、しばらくは固定する必要がある。

大学教育充実の一環として、国試の模試とも言われている総合卒試(総合卒業試験)の問の質の向上に努め、これを用いて国試の可否を高い確率で予測できれば、不合格となる可能性の高い者にもう1年間在学のまま勉強する機会を与えることによって、ストレート合格率をあまり下げずに、世間一般で使われている国試合格者を飛躍的に高め、質の高い医師を社会に送り出すことが可能になる。また、卒延になった学生はそのことを納得し、学生全体に勉学意欲が高まれば、正に一石三鳥である。

そこで、総合卒試の各問の識別性能を点検し、識別性能の高い問だけによる得点を用いて国試の可否を予測するためのデータ解析方法並びにコンピュータ・ソフトを開発し、改良を加えてきた<sup>5)6)</sup>。今後ともソフトの改良に努めたいと思っているが、もし同様なテストが行われている全国の大学の協力が得られれば、一度出題された問の中から主として識別性能の高いものを選び、コンピュータに蓄積することによって、その比率を60%としても、10年間で1回の試験に必要な問の480倍(0.6×10×80大学)の問が集まることになるから、すでに米国で行われているような、学生が随時コンピュータの前に座ってテストを受けることも夢ではなくなる<sup>7)</sup>。また、その際は問を正答率によって平易、普通、難問等に分け、テストを初級、中級、上級とすることもできる。

## 2 卒業判定と国試合格率の関係

本学には「必修科目が1科目でも不合格の場合は留年・卒延」という規則があるが、各科目の実質的な合格基準はまちまちである。91年(91年3月卒業予定の者が対象という意味、以下同様)までは、留年者も比較的少数で、卒延者は6年間で4人である。92年から卒業判定が

多少厳しくなったように思えたが、国試合格率の高い年が続くと甘くなり、低い年が続くと厳しくなる傾向のあることを否定できない。そして、それまでは2年、4年、6年終了時に行われていた進級判定が、92年入学者(98年)から各学年で行われるようになり、6年次に進級するまでに、従前に比べて大量の留年者が出たため、98年の総合卒試受験者は過去最少より11人も少ない81人で、卒延者は1人だけであった。もし、6年次までは従前のように進級させていたら、もっと多数の卒延者が生じていたものと思われる。

そこで、98年の卒延者を過去最多に1を加えた6人とし、受験者を86人に補正したものについても修正合格者数及び相関係数を求め、表1及び表2に示した。ちなみに、98年の受験者81人中92年入学者は79人である。なお、ここではその年に総合卒試を受けた者を対象としているので、データの中には入学した年が異なる者も含まれている。また、97年までは国試の模試的傾向のあった総合卒試が、98年には卒業判定の有力な資料となっている。ここで、表1及び2の各変数は次のように定義され、図3はこれらの値(98年は補正值による)を91年以前と92年以降を別々に標準化して描いたものである。

受験者(a): 総合卒試受験者(卒業予定者)数

卒延者(b): 受験者の中で卒業できなかった者の数

表1 卒延者数、不合格者数等の推移

卒年	受験者	卒延者	不合格者	全 国 合格率	合格率	修正合格者数
86	92	1	6	86.6	93.3	92.3
87	105	2	5	86.2	95.2	93.4
88	100	0	10	81.2	92.7	92.7
89	96	0	12	88.0	85.8	85.8
90	104	1	13	82.9	89.9	89.1
91	98	0	11	84.3	90.2	90.2
92	97	5	4	84.0	96.3	91.3
93	106	4	4	90.1	94.6	91.0
94	100	0	17	86.2	83.2	83.2
95	96	1	11	86.0	88.7	87.8
96	96	4	3	89.3	95.8	91.9
97	105	0	16	88.1	82.5	82.5
98	81	1	1	89.6	98.4	97.1
	(86)	(6)				(91.5)
mean	98.2	1.5	8.7	86.3	91.3	89.9
	(98.5)	(1.8)				(89.4)
s. d.	6.7	1.8	5.2	2.7	5.1	4.2
	(5.7)	(2.2)				(3.6)

( )は補正值による

表2 卒延者数、不合格者数等の間の相関係数

	卒 延 者	不 合 格 者	合 格 率	修正合格者
卒 延 者 数	-0.697	0.646	0.383	
不 合 格 者 数	-0.858	-0.954	-0.874	
合 格 率	0.798	-0.954	0.952	
修正合格者数	0.502	-0.817	0.920	

対角線の右上は98年卒の受験者数81、卒延者数1、左下は同受験者数86、卒延者数6として求めたもの  
 $\rho=0$ の仮説の下で  $n=13$  のとき、 $P(r>0.476)=0.05$ 、 $P(r>0.553)=0.025$ 、 $P(r>0.684)=0.005$

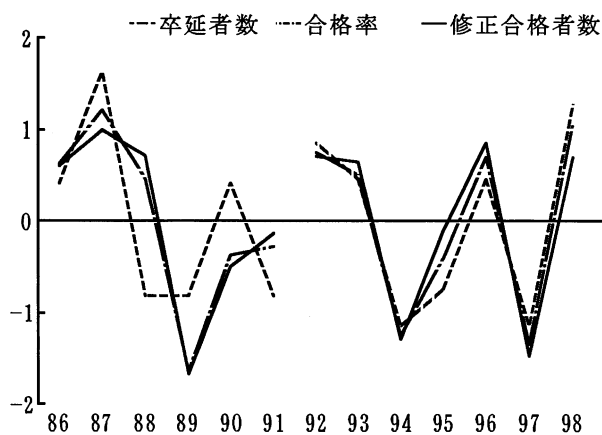


図3 卒延者数, 合格率, 修正合格者数の推移

不合格者(c): 受験者の中で卒業し, 国試に不合格となった者の数

$$\text{合格率} : P = \left( 1 - \frac{c}{a-b} \times \frac{100-h}{100-g} \right) \times 100 \text{ (\%)}$$

不合格率を全国不合格率を考慮して補正し, 100から引いたもの。但し, gは全国合格率, hは全国合格率の13年間の平均

$$\text{修正合格者数} : x = \left\{ a - \left( b + c \times \frac{100-h}{100-g} \right) \right\} \times \frac{100}{a}$$

全国不合格率を考慮して不合格者数を補正したものに卒延者数を加えたものを受験者数から引き, 受験者数が100になるように補正したもの。

さて, 卒業判定を厳しくしたとき, 国試の合格率が上がるのは当然として, 卒延者を不合格者とみなした修正合格者数が増えていることが注目される。Pとxの相関係数(r)が0.95(補正值を用いると0.920, 以下同様)と大きいのは当然として, bとPのrも0.64(0.798)と大きく, bとxのrは0.38(0.502)である。そして, 92年以降では, Pとxのrが0.975, bとPのrが0.977, bとxのrが0.911で, b, P, xのグラフがほぼ完全に一致している。なお, P(母相関係数)=0の仮説の下に, n=7のとき P(r>0.875)=0.005である。このように, 大学審議会の報告を待つまでもなく, 卒業判定を甘くすることは, 学生のためにならないことが分かる。

### 3 採点・集計・統計・追跡用プログラム

総合卒試の問の解答は多肢選択(五者択一)方式で, 答案にはマークシートを用いる。なお, 96年まではA問題~E問題の中に合せて327問あったが, 97年に法医学と救急医学が加わって, 表3のような335問に増えた。そして, 98年にはF問題ができ, 内科の問が大幅に増えたが, 問数の合計は326に減った。

表3 総合卒業試験の科別・試験(問題)別の問の数(左97, 右98)

	A 問題		B 問題		C 問題		D 問題		E 問題		F 問題	合計	
内科学 1	7	8	10	15	3	3	5	4	5	5	5	30	40
内科学 2	7	8	10	15	3	3	5	4	5	5	5	30	40
内科学 3	7	8	10	15	3	3	5	4	5	5	5	30	40
外科学 1	7	3	6	2	3	3	5	2	5	5	5	26	20
外科学 2	7	3	6	2	3	3	5	2	5	5	5	26	20
小児科学	10	5	12	9	3	3	5	4	5	5	5	35	31
産婦人科学	10	5	12	10	3	3	5	2	5	5	5	35	30
保健学	4	10	6				2	1	2	2	2	14	15
保健学	4	10	6				2	1	2	2	2	14	15
精神神経医学	4	2	2	2			1	1	1	1	1	8	7
耳鼻咽喉科学	4	2	2	2			1		1	1	1	8	6
皮膚科学	4	2	2	2			1		1	1	1	8	6
整形外科学	4	2	2	2			1		1	1	1	8	6
脳外科学	4	2	2	2			1		1	1	1	8	6
麻酔科学	4	2	2	2			1		1	1	1	8	6
泌尿器科学	4	2	2	2			1		1	1	1	8	6
眼科学	4	2	2	2			1	1	1	1	1	8	7
放射線医学	4	2	2	2			1		1	1	1	8	6
歯科口腔外科学	4	2	2	2			1		1	1	1	8	6
臨床検査医学	7	7										7	7
法医学	3						2					3	2
救急医学	3						2	2	2			5	4
合計	116	87	98	88	21	21	51	30	49	51	49	335	326

プログラム 1 採点・集計用

- (1) 問題( A ~ E , 98年はFまで)別に答案を 2 回ずつ読む。
- (2) 正答か, 誤答か, 判別できない(マークが悪い)かを調べ, 1 回目と 2 回目的一致しているか否かをチェックする。「判別不能」又は「不一致」のものがある場合, 何枚目の答案のどの問かを印刷する。その答案を調べ, 必要ならマークを修正する。
- (3) 問別に正答者数及び正答率を求める。正答率の極めて低い問については, 直ちに出题者に報告する。その結果, 正解が訂正されることもあるが, その場合はその問題について(1)からやり直す。
- (4) 問題別の個人得点及びその和(総合点)を求め, 受験番号順に並べる。
- (5) 問題別及び全体(総合点)の問数, 平均, 標準偏差, 正答率, 100点換算の標準偏差, 点数の分布を求め, 分布のグラフを描く。
- (6) 問題別得点及び総合点について, 個人の偏差値, 順位, 得点率, 科別得点を求め, 総合点順位によって並べる。また, 科別の問数, 平均, 標準偏差, 正答率, 100点換算の標準偏差を求める。
- (7) 問題間の相関行列及び科間の相関行列を求める。
- (8) (6)で求めた個人成績に(5)及び(6)で求めた問題別及び科別の平均, 標準偏差等を付けて印刷する。

そこで, (8)の結果を本人に渡すことになるが, 性能のよいマークシート・リーダーを用いれば, 試験終了後 1 時間以内でここまで終えることができる。なお, 98年から禁忌肢がある問が入ったので, 各受験者の禁忌肢選択状況及び各禁忌肢の選択者一覧を印刷するようにした。

プログラム 1 を用いると, (1)~(8)で述べたことのほかにも, いくつかの興味ある結果が得られるが,  $x_{ik}, x_{jk}$  が標準化された値をとるとき, 相関行列  $R = (r_{ij})$  から

$$d_{ij} = \sqrt{\frac{1}{n} \sum_{k=1}^n (x_{ik} - x_{jk})^2} = \sqrt{2(1 - r_{ij})}$$

によって距離が求まるから, クラスター分析によって<sup>8)</sup>, A ~ F ( E )あるいは科をグループに分類することができる。群平均法によるクラスター分析を行うと, 図 4 及び図 5 のようなデンドログラムを得る。図 4 のデンドログ

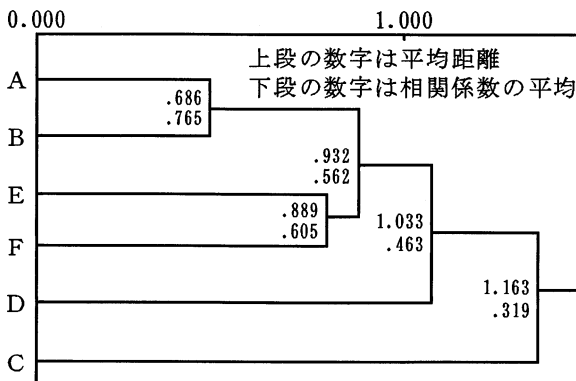


図 4 問題 A ~ F の群平均法によるクラスター分析を用いて得られたデンドログラム ( 98 )

ラムは表 3 から予想されるもので, 毎年同じような傾向がみられるが, 97年は平均点が低く上位者と下位者の相違が明確である上に問数も多いので, 98年よりかなり距離が短い。図 5 については「問数の多い科同士の距離が短くなる」という傾向がみられるだけで, それらの科を除くと, 結び付きはあまり強くない上に年によって異なる。すなわち, 1 つの科の出す問数がこの程度では, 科の類似性はほとんどみられない。

プログラム 2 統計・追跡用

- (1) 総合点によって受験者を最上位( H ), 上位( HM ), 中位( M ), 下位( LM ), 最下位( L )の 5 群に分ける<sup>1)</sup>。そして, L を仮不合格者群, L 以外の 4 群を合せたものを仮合格者群 T とする。次に, 上位( A ), 中位( B ), 下位( C )の 3 群に分ける。また, 国試の合格者

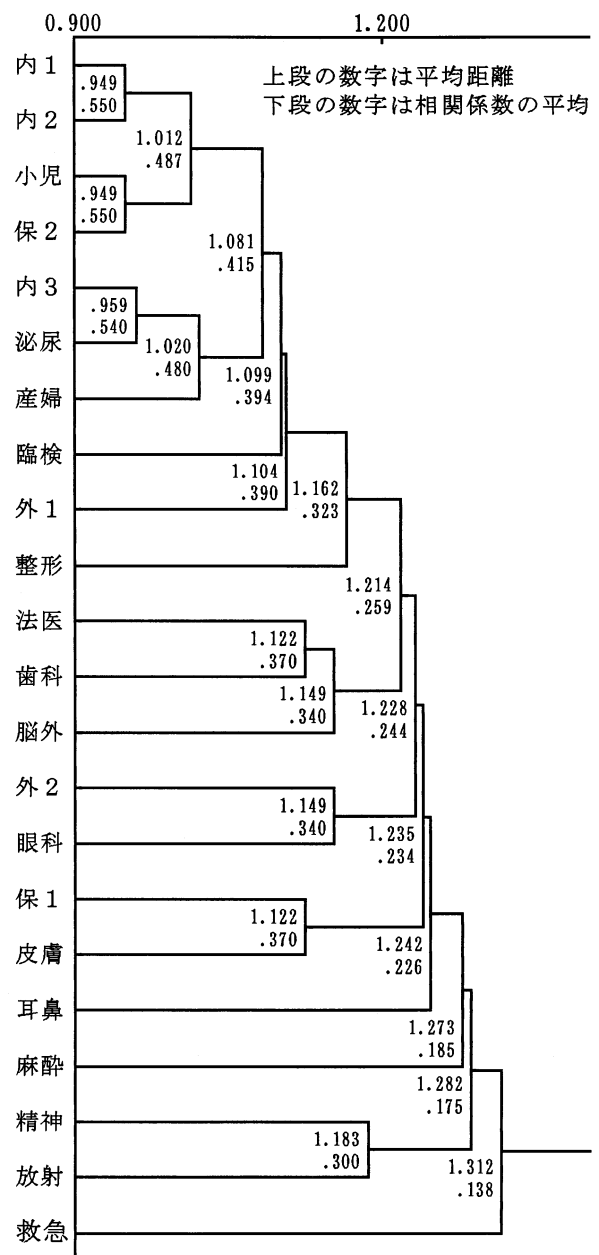


図 5 各科の群平均法によるクラスター分析を用いて得られたデンドログラム ( 98 )

が判明した時点で、国試合格者(S), 国試不合格者(F)の2群に分ける。

以下(6)まで、各問について、

- (2) (1)の各群及び全体の正答率を求める。
- (3) 5群に関して、下から  $i = 1, 2, \dots, 5$  の各群の正答率  $x_i$  の間で、「 $i < j$  かつ  $x_i > x_j$ 」が成り立っているとき、言いかえると、総合点の低い者のグループの正答率が高い者のグループの正答率を上回っているとき、転位が起きているとし、すべての  $i, j (i \neq j)$  の組合せに対する転位の数(転位数と呼ぶ)を求める。また、3群について、正答率の大きい方から並べる。
- (4) 正答を1, 誤答を0として、総合点との間の相関係数( $r$ )を求める。
- (5) 5群に分けた場合の正答率曲線を描き、好ましくない( $r$ が負、転位数が大きい、Lの正答率が高いなど)場合には出題者に報告する。(正解が訂正される可能性もある)
- (6)  $p_1, p_2$ をそれぞれ仮合格者群, 仮不合格者群の正答率,  $p$ を全体の正答率,  $n_1, n_2$ を両群それぞれの人数とすると、仮合否両群の母集団正答率に差がないとして、両群の正答率の差を標準化した値

$$u = \frac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

を求める。また、国試の合否が判明したら、同様な仮説の下で、合否両群の正答率の差を標準化した値( $z$ )を求める。

- (7) 正答率( $t\%$ ),  $z, u, r$ , 転位数( $tr$ ), 3群の順序( $od$ )後の3つは本質的に同じ)のうち2変数の間

の散布図を描き、階級ごとの平均, 標準偏差を求め、関連の強さを知るために $\chi^2$ 及びクレーマーの関連係数を求める。また,  $t, z, u, r$ の間では相関係数も求める<sup>5)</sup>。

- (8)  $r, u$ の大きさ,  $tr, od$ に関する条件を設け、その条件を満たす問すなわち識別性能の高い問だけによる得点及び順位を求め、国試不合格者がいかに下位に集中するか調べる<sup>5)</sup>。また、合否両群の母集団に差がないという仮説の下に、不合格者の平均順位を標準化したもの( $z_1$ ), この得点の合否両群の平均値の差を標準化したもの( $z_2$ )を求める。
- (9) 特に,  $r$ 及び $u$ に関する条件の下限の値を連続的に変化させて、(8)と同様のことを行い、どの辺りのとき国試不合格者が最も下位に集中するか追究する。

#### 4 総合卒試の各問の実力識別性能

表4に示したように、96年以降は、不合格者が、5群に分けた場合にはL, 3群に分けた場合にはCに集中している。また、5群に分けた場合に、不等式  $0 < x_1 < x_2 < \dots < x_5 < 1$  をすべて満たす問は、97年に最高の20%である。しかし、転位数の大きい問の比率が最も低いのは追跡調査を始めた最初の94年である。この年は国試不合格者が史上最多の17人で、上位者と下位者の実力の差が大きかったのであろう。98年には転位数の大きい問の比率がやや高くなっているが、各群の人数が少ないこと及び全員正答の問以外にも正答率の高い問が多いためグループ間の差が判別し難くなっていることが原因として考えられる。

表4 年度別の各グループの人数

卒年	受験者数	グループ分け1 (5群)					グループ分け2 (3群)			グループ分け3 (2群)		国試不合格者数
		L	LM	M	HM	H	C	B	A	F	S	
94	100	20(9)	20(5)	20(0)	20(2)	20(1)	32(13)	34(2)	34(2)	17	83	17
95	96	19(6)	19(5)	20(1)	19(0)	19(0)	31(10)	33(2)	32(0)	12	84	11
96	96	20(6)	18(1)	20(0)	17(0)	21(0)	32(7)	31(0)	33(0)	7	89	3
97	105	21(14)	20(1)	22(1)	21(0)	21(0)	35(15)	35(1)	35(0)	16	89	16
98	81	16(2)	17(0)	16(0)	15(0)	17(0)	26(2)	28(0)	27(0)	2	79	1

L: 最下位, LM: 下位, M: 中位, HM: 上位, H: 最上位, C: 下位, B: 中位, A: 上位, F: 不合格者(卒延者を含む), S: 合格者, ( )内は不合格者の内数

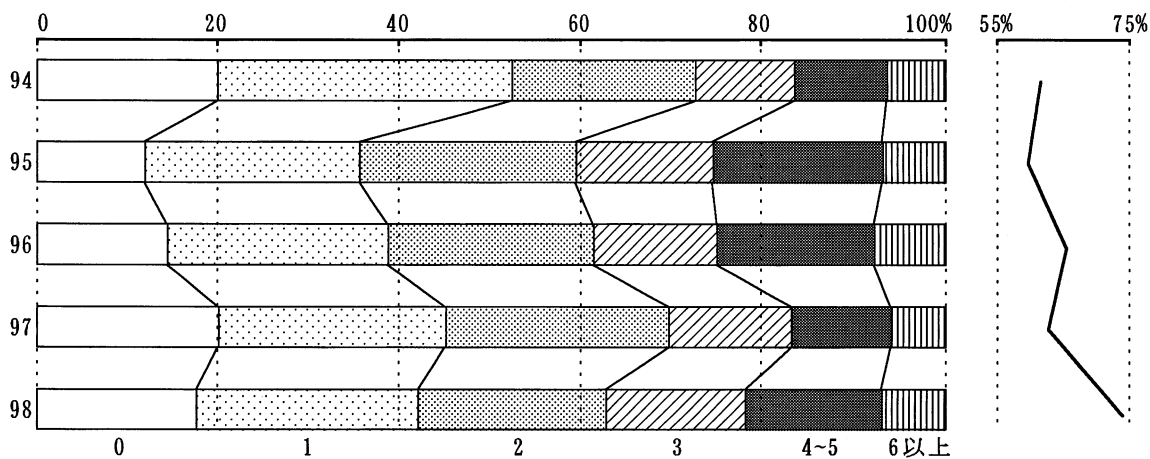
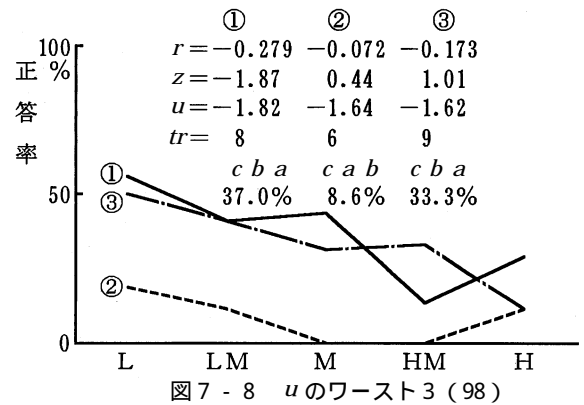
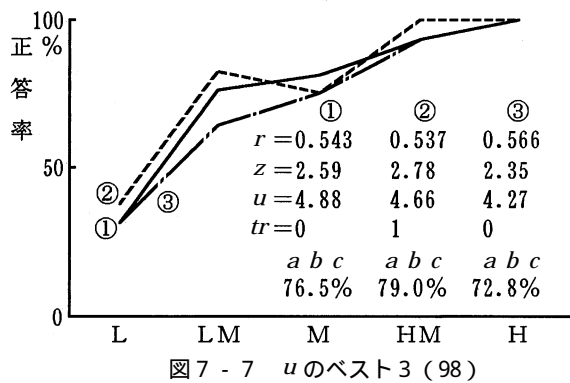
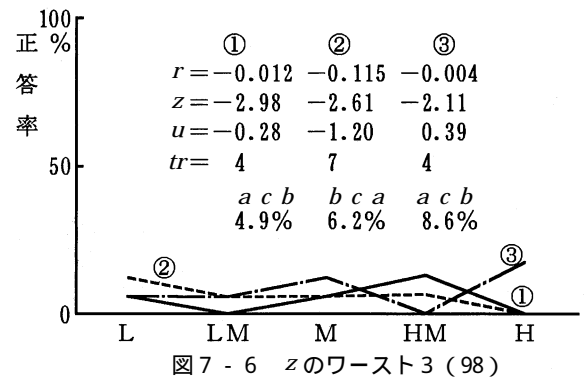
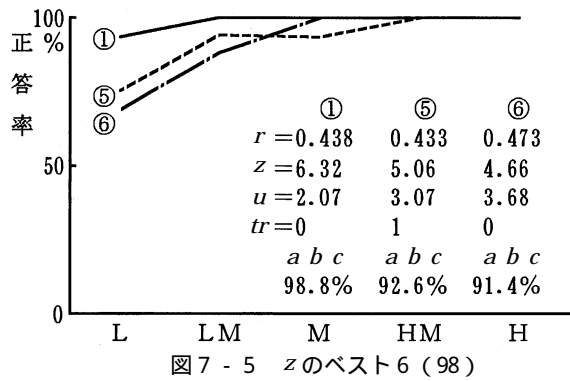
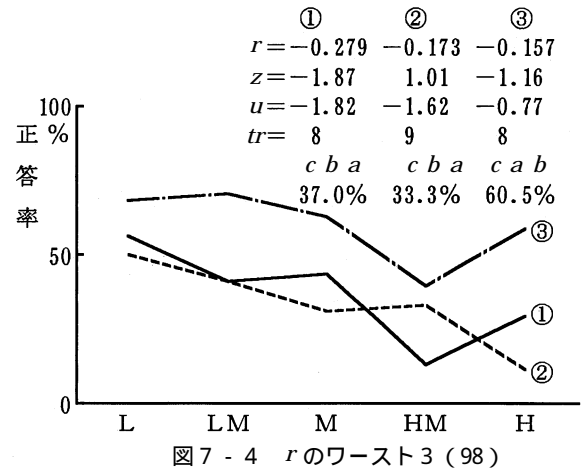
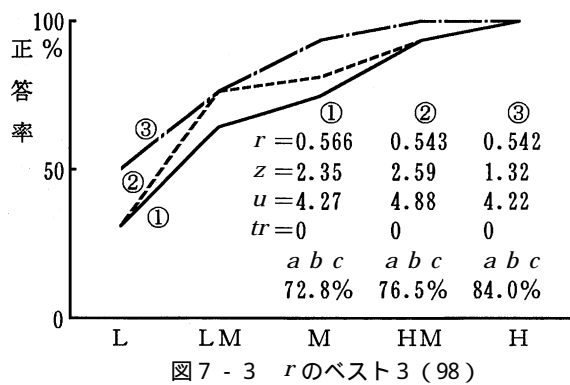
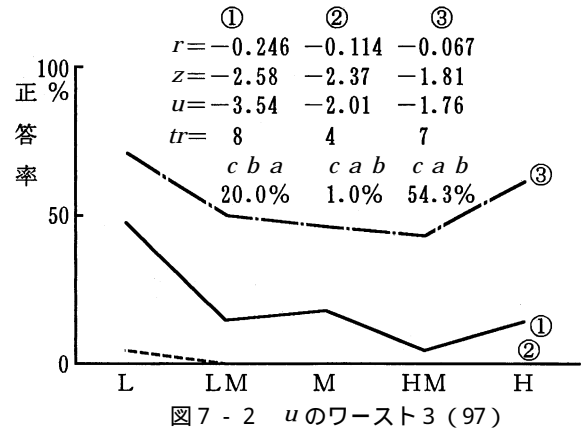
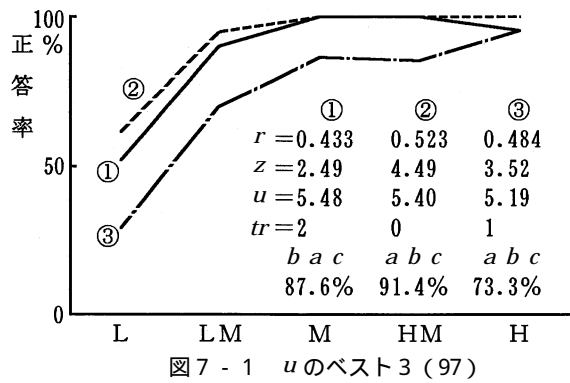


図6 転位数別の問の比率 (正答率100%及び0%の問を除く)

平均(全問)正答率



$r, z$  のベスト3及びワースト3について、正答率曲線を図7に示した。なお、97年については文献5)に示してある。 $r$  のベスト3の間は、いずれも  $tr=0$  で、 $z$  も大体有意な値になっている。これに対して、 $r$  のワースト3の間は、いずれも  $tr=8\sim 9$  で、 $t$  もかなり低いものが多い。しかし、中には  $t=70$  でありながら、 $r=-0.228, z=-1.77$  という実力識別性能の全くない間もある<sup>5)</sup>。「総点の高い者が実力がある」と考えれば、「 $r$  が大きい間は実力識別性能が高い」ということになるが、国試の合格者と不合格者を識別する性能は  $z$  の大きさに計られる。ところが、 $z$  は不合格者が分かった後に算出されるものであるから、合否予測にはこれと相関の高いものを代りに用いねばならない。

不合格者の多い年は一般に  $z$  も大きくなるので、異なる年の間で問の識別性能を単純に  $z$  の大きさに比較するのはあまり適当ではないが、97年及び98年は合格者と不合格者を明確に識別するという点で優れた問が多く、 $z$  のベスト3は、いずれも  $5.0 < z$  を満たし、正答率は80%以上である。また、いずれもLMとLの正答率の差が大きい。このことから、「L群に属する者を仮の不合格者として、 $z$  と同様のもの( $u$ )を求め、これを用いて合否予測を行えば、 $r$  による場合よりさらに予測力が高まるのではないか」ということを思い付き、 $u$  のベスト3、ワースト3の正答率曲線も図7に加えた。なお、 $z$  のワースト3は、いずれも  $z < -2.0$  を満たし、正答率が20%以下、特に98年は9%を割っている。

97年は、 $u$  の2番目に大きい問が  $r$  の最大の問と同じ問であるが、全体的にみて、 $u$  のベスト3の方が  $r$  のベスト3より  $z$  のベスト3と正答率曲線が類似している。また、 $u$  のワースト2は  $z$  のワースト2と一致している。しかし、98年は不合格者が2人だけのため  $z$  のベスト4は全く同じ曲線で、 $t=98.8$  と非常に高く、 $u$  のベスト3、ワースト3の正答率曲線が  $r$  のそれと類似し、 $z$  のそれとはやや趣を異にしている。

$r$  と  $z$  と  $u$  の間の相関係数の年次変化をみると、 $r$  と  $u$  の間が大きいのは当然として、 $r$  と  $z$ 、 $u$  と  $z$  の間も年を経るごとに大きくなり、97年には飛躍的に大きくなっている。98年にはいずれも小さくなっているが、不合格者が2人だけであるからやむを得まい。

以上のことから、国試の合否予測には  $r$  あるいは  $u$  の小さい問を除外して求めた点数を用いれば、事足りると

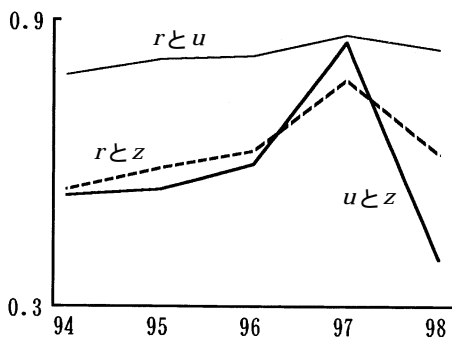


図8  $r, z, u$  の間の相関係数の推移

考えられるが、 $r$  や  $u$  の値だけでは視覚に訴える力が弱い上に、一般の人に相関係数を理解してもらうことは相当困難であるから、正答率曲線を描いてみるのが有効である。また、「その間では、総点の低い者のグループの正答率が、総点の高い者のグループの正答率を上回っている」と言えば、誰にでも理解され得るから、転位数も、一般にCの位置で識別性能がおよそ分かるA, B, Cの正答率の順序とともに、個々の問の識別性能を見当付けるのに役立つ。

### 5 総合卒試による国試の合否予測

$0.02(i-1) < r$  という条件を満たす問だけによる得点を用いて、合否予測をすることにし、 $i$  の値を変化させると、3節で定義した  $z_1, z_2$  がどのように変わるだろうか。 $z_1$  より変化がはっきり読める  $z_2$  についてそのグラフを図9に示した。また、 $0.18(i-1) < u$  という条件を満たす問だけによる得点に関する  $z_2$  のグラフを図10に示した。

97年は  $r$  と  $z$  及び  $u$  と  $z$  の相関が高い上に不合格者も多いので、 $z_2$  は他の年に比べ際立って大きい。 $r$  の値によって問を制限した場合、94年~96年は  $i=15\sim 16$  すなわち  $0.28\sim 0.30 < r$  辺りに  $z_2$  のピークがみられるが、97年は制限を厳しくするに従ってほぼ直線的に  $0.36 < r$  (これを満たす問の数が58, 以下同様) まで増大し、 $0.38 < r(49)$  の場合にも減少幅は微小である。また、 $u$  の値による場合、97年は  $3.06 < u(68)$  にピー

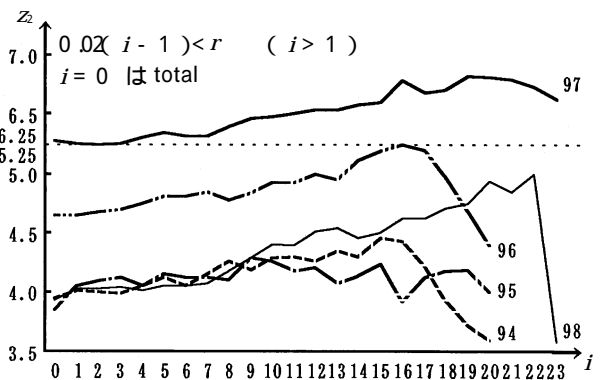


図9  $r$  の大きさによって問を制限した場合の  $z_2$  の変化

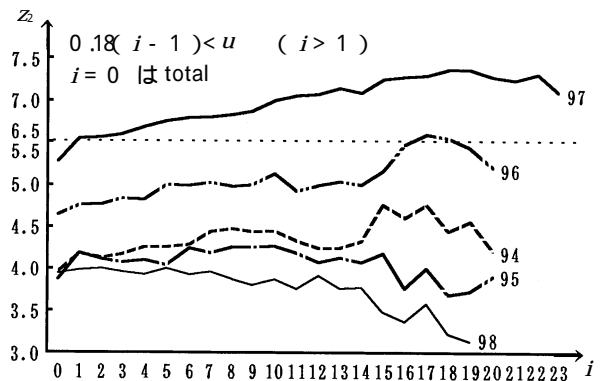


図10  $u$  の大きさによって問を制限した場合の  $z_2$  の変化

クがあり、 $r$ による場合よりやや大きくなっている。98年は不合格者が2人だけであるためか、 $r$ による場合は $0.42 < r(21)$ まで増加しているが、 $u$ による場合は $0.18 < u(238)$ にピークがある。

94年～96年の場合は、 $0.3 < r$ を満たす問による得点で順位を付ければ、下から10番までの内で不合格者の占める比率が50%程度から70%程度に上昇する。しかし、97年は総点でも下から10番の内の9人が不合格者であるから、 $r$ や $u$ の値によって問を制限すること自体は、国試の合否予測力を高めることにあまり寄与していない<sup>5)</sup>。また、98年はどのように得点の計算方式を変えても、卒業者(1人)は最下位であるが、国試の不合格者(1人)が下から2番目とはならない。98年は不合格者が非常に少なかったことを評価すべきであろう。

## 謝 辞

本研究は、平成7年から3年間にわたる科学研究費基盤研究(A)(多変量データ解析の利用による大学入試データ解析システムの開発)の一環として行ったもので、代表者の柳井晴夫大学入試センター研究開発部長には、大変お世話になった。古川進山梨大学工学部教授には、コンピュータ・ソフト、とりわけグラフィック・ソフト開発の面で多大なご協力をいただいた。本学在職中は入試の追跡調査・研究における共同研究者であった恵泉女学園学園長の川田殖先生には、原稿に目を通していただき、貴重なご意見をいただいた。入学者選抜方法研究委員会研究補助員の秋山友紀さんには、データの整理、コンピュータへの入力、ワープロによる原稿作成の

一切を担当していただいた。以上の方々に、日ごろのご支援と合せて、深く感謝の意を表したい。

## 文 献

- 1) 高野文彦:(1992)試験の評価方法としての項目反応の応用, 大学入試研究ジャーナル, 第2号, 1-13
- 2) 平野光昭:(1994)医師国家試験の大学としての成績を高める入試及び他の要因 主成分分析。大学入試研究ジャーナル, 第4号, 6-13
- 3) 平野光昭:(1995)入試成績・入学時の属性・学内成績と医師国家試験の合否の関係。大学入試研究ジャーナル, 第5号, 39-49
- 4) 平野光昭, 渋谷昌三:(1996)高校調査書に記載された成績及び諸活動と医師国家試験の合否の関係。大学入試研究ジャーナル, 第6号, 76-83
- 5) 平野光昭:(1997)医師国家試験の合格率を高めるために 入学者選抜・大学教育・総合卒業試験。山梨医科大学紀要, 第14巻, 50-60
- 6) 平野光昭:(1998)総合卒業試験の改善並びに医師国家試験合否予測のためのデータ解析。多変量データ解析の利用による大学入試データ解析システムの開発(平成7,8,9年度科学研究費補助金基盤研究(A))研究成果報告書, 377-384
- 7) 池田 央:(1998)米国にみる試験方法の技術革新。国立大学入学者選抜研究連絡協議会第19回大会セミナー資料, 18-28
- 8) B.F.J.マンリー著, 村上正康・田栗正章共訳:多変量解析の基礎, 培風館

## Abstract

### On the Development of the System of Data Analysis for the sake of Enhancing of the Discriminating Power of the Comprehensive Graduation Test (CGT) and Accelerating Its Technical Innovations

Teruaki HIRANO

The reform of the entrance examination may be called a success provided that it results in the increase of the number of medical students who are to go straight on to the doctors in due course of six years and in the birth of promising doctors full of human excellence. However, as the pass-rate of national examination for medical license (NE) is often considered as a landmark in the evaluation of medical education of respective universities, we cannot disregard the elevation of the pass-rate of NE. Therefore, we developed the system of data analysis and endeavored to improve it for the past three years both to examine the powers of discrimination of questions in CGT which has close affinities with NE and, using the particularly discriminated questions, to predict the pass-rate for any student. In this essay, the author introduces what kinds of analytic methods and computer programs were developed, and, applying these methods, presents the results of analyses of the data of CGT for the students who graduated between '94 and '98. In case we can obtain cooperation from many universities where the same tests are given as CGT, by storing in computers a great many questions which seem to serve as the decisive factors to distinguish between the able student and the unable one, it might be no longer a dream that any medical student sits down to the computer and takes the test, as already done in U. S. A.