

## 学内成績と医師国家試験の合否の関係 - 特に総合卒業試験による合否予測 -

平野光昭

入学する学生の質を高め、教育の質を向上させることは、6年間でストレートに医師となる者の数を増やすとともに、人間性豊かな将来性のある医師を誕生させることになるが、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)の間の識別性能を点検するため、かつ国試の合否を予測するためのデータ解析システムを開発し、数年間にわたり改良を加えてきた。本論文では、新たに改良した方法を適用して、1999年卒及び2000年卒の者を対象に行われた総合卒試のデータを解析し、学内成績と国試の合否の関係についても考察する。もし、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，学内成績，合否予測

### 1 はじめに

86年から2000年(以下、00年と言う。)までに1500人近い医師を誕生させた本学では、表1に示したように、95年まで毎年90人以上(入学者100人)がストレート(6年間)に卒業し、97年までストレートに国試(医師国家試験)にも合格して医師となった者の比率が一度も80%を割ったことがなかった。しかし、国試の大学別合格率の順位が常に上位にあったわけではなく、94, 95, 97年には下に甘んじた。98年には、全国合格率を考慮し卒延者を不合格者と見なして算出した修正合格者数が、97.1人という最多の値となったが、ストレート合格者数は、98年から毎年最少値を更新し、94年に入学した者は3分の2しか6年間で医師になれなかった。修正値も71.1人と15年間で最少である。

単なる合格率より全国の大学の中でのその順位が目されるのは、毎年全国合格率が変動しているからで、全国的に見て、合格者を10%も上下させる程、年によって受験者の質が異なるとは考えられず、その違いは主として判定(問題)の難易によるものと考えられる。また、一般に卒業判定を厳しくすれば合格率が高まるので、入学試験及び大学教育の評価の物差しとして、ストレート合格者数が注目される。そこで、年度間を比較するものとして、全国の不合格率が大体10%~20%と低い値であることから、学生の質が年度によって変らなければ、本学の不合格率が全国の値に比例するものと考え、次のような修正値を用いることにする<sup>4)~7)</sup>。

国試受験者数を  $a$ 、そのうちの不合格者数を  $b$ 、全国

合格者を  $c$  %、全国合格率の15年間の平均値を  $d$  % とするとき、修正合格率 ( $p$  %) は

$$p = \left( 1 - \frac{b}{a} \times \frac{100-d}{100-c} \right) \times 100$$

また、修正ストレート合格者数  $m$  は、ストレート卒業生数を  $e$ 、そのうちの不合格者数を  $f$  として、

$$m = e - f \times \frac{100-d}{100-c}$$

さらに、修正合格者数  $n$  は、卒延者数を  $g$  として、

$$n = \left( a - b \times \frac{100-d}{100-c} \right) \times \frac{100}{a+g}$$

すなわち、総合卒試(総合卒業試験)受験者(卒業予定者)を100人とし、全国合格率が例年並ならば、そのうちの何人が国試に合格したかという数が  $n$  である。

大学審議会の「21世紀の大学と改革の方策について」と題する答申でも、卒業認定を厳しくすることがうたわれているが、大学教育充実の一環として、99年卒業予定者から、従来国試の模試と言われていた総合卒試を卒業判定の主たる資料として使うことにした。しかし、卒業を前にして卒延になった者は99年が0人で、00年もわずかに1人である。ところが、表1及び2に示したように、卒延者数と修正合格者数の間の相関係数は正である。もともと本学の場合卒延者は毎年わずかであるが、その数が比較的多い5, 4, 4である92, 93, 96年は、不合格者数が4, 4, 3で、卒延者を不合格者と見なしても他の年に比べて少ない。

総合卒試の質の向上に一層努め、これを用いて国試の合否を高い確率で予測できれば、不合格になる可能性の高い者にさらに1年間在学のまま国試の勉強をする機会

表1 ストレート卒業者数, 卒延者数, 不合格者数等の推移

卒年	ストレート 卒業者	ストレート 合格者	総合卒試 受験者	卒業時 卒延者	国 試 不合格者	全国合格率	修正ストレート 卒業者合格率	修正合格率	修正ストレート 合格者	修正合格者
86	90*	84	92	1	6	86.6	92.9	93.0	83.6	92.0
87	97	94	105	2	5	86.2	96.8	95.0	93.9	93.2
88	96	88	100	0	10	81.2	93.7	92.4	89.9	92.4
89	91*	80	96	0	12	88.0	85.6	85.1	77.9	85.1
90	97	86	104	1	13	82.9	90.5	89.5	87.8	88.6
91	93*	85	98	0	11	84.3	92.2	89.8	85.7	89.8
92	90	86	97	5	4	84.0	96.0	96.1	86.4	91.2
93	92	90	106	4	4	90.1	96.9	94.3	89.1	90.8
94	93	80	100	0	17	86.2	85.5	82.4	79.5	82.4
95	93	82	96	1	11	86.0	87.9	88.2	81.8	87.3
96	87	85	96	4	3	89.3	96.9	95.6	84.3	91.7
97	93	82	105	0	16	88.1	85.8	81.7	79.8	81.7
98	78	77	81	1	1	89.6	98.2	98.3	76.6	97.1
99	86	76	104	0	11	84.1	89.6	90.5	77.0	90.5
00	80	67	89	1	14	79.1	88.9	89.1	71.1	88.1
mean	90.4	82.8	97.9	1.3	9.2	85.7	91.8	90.7	83.0	89.4
s.d.	5.6	6.4	6.9	1.7	5.0	3.2	4.5	4.9	6.1	4.1

\*ストレートは沖縄留学生を除く100人が対象。

表2 ストレート卒業者数, 卒延者数, 不合格者数等の間の相関係数

	S 卒業者	S 合格者	受験者	卒延者	不合格者	修正S合格率	修正合格率	修正S合格者	修正合格者
S 卒業者	1.000	0.760	0.767	-0.076	0.282	-0.172	-0.305	0.751	-0.320
S 合格者	0.760	1.000	0.559	0.383	-0.388	0.471	0.309	0.969	0.237
受験者	0.767	0.559	1.000	0.039	0.314	-0.186	-0.322	0.549	-0.383
卒延者	-0.076	0.383	0.039	1.000	-0.690	0.669	0.636	0.387	0.376
不合格者	0.282	-0.388	0.314	-0.690	1.000	-0.925	-0.935	-0.371	-0.854
修正S合格率	-0.172	0.471	-0.186	0.669	-0.925	1.000	0.957	0.520	0.892
修正合格率	-0.305	0.309	-0.322	0.636	-0.935	0.957	1.000	0.379	0.954
修正S合格者	0.751	0.969	0.549	0.387	-0.371	0.520	0.379	1.000	0.320
修正合格者	-0.320	0.237	-0.383	0.376	-0.854	0.892	0.954	0.320	1.000

S 卒業者, S 合格者等については表1を参照.  $\rho = 0$  の仮説の下で,  $n = 15$  のとき,  $P(r > 0.441) = 0.05$ ,  $P(r > 0.514) = 0.025$ .

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, 国試合格率の大学間順位を飛躍的に高めることが出来る。また, 卒延になった学生に納得する理由を示すことにもなり, 後に続く学生の学習意欲が高まれば, 正に一石三鳥である。

一方, 幅広い教養と優れた問題解決能力を身に付け, 患者から信頼される人間性豊かな医師を社会に送り出すには, 教養科目を初め, 国試の出題範囲外の科目をもしっかり勉強してもらわないと出来ないが, 低学年で進級判定を厳しくすることは, 国試そのものが改善されない限り, ストレート卒業者を減少させるが, 合格率の大学間順位の向上にはあまり効果がない<sup>1)~4) 6)</sup>。ちなみに, 92年入学者から, 6年次までの進級判定が厳しくなったので, ストレート合格者数が減少しているが, ストレート卒業者の合格率はそれ程高まっていない。もちろん, 医学教育の評価基準の1つとして, 国試の大学としての成績が世間一般から注目されているが, 医学教育の目標が国試の合格率を高めることだけにあるわけではない。

ともあれ, 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を追究してきた<sup>5)~7)</sup>。当初解答は単純な五者択一方式であったが, 98年から禁忌肢, 00年からは複数解答(2つの選択肢を選び, 両方合っている場合に正答となる。)が加わり, それらに関係してコンピュータ・ソフトを改良した。しかし, 00年は国試不合格者14人(他に卒延者1人)のうち, 総合卒試で最下位

グループ(下から5分の1)に属する者が半分の7人だけという異変が起きている。この原因を追究するため, 総合卒試以外の6年次の成績及び1年次から5年次までの成績等と国試の可否の関係についても考察する。

なお, 将来は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, すでに米国で行われているように, 学生が希望する時, コンピュータの前に座って総合卒試が受けられるようにすることを目指している。

## 2 総合卒業試験の各問の実力識別性能

内科学3の一部が神経内科となった外は, 98年以来総合卒試の問題(A~F)別・科別の問数は変っていない<sup>6)</sup>。総点によって, 受験者を最上位(H), 上位(HM), 中位(M), 下位(LM), 最下位(L)の5群にほぼ等分し, 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群にほぼ等分する。また, 国試の可否が判明した時点で, 国試合格者群(S)と国試不合格者群(卒延者を含む)(F)の2群に分ける。そして, 各問について, 各群及び全体の正答率を求める。

表3に各群の人数とそのうちの不合格者の数を示した。Lに属する不合格者数とTに属する不合格者数が, 94年と95年はほぼ同数であるが, 96年から98年までは不合格者のほぼ全員がLに属し, HM以上に属する不合格

表3 総合卒試の成績等によって分けられた各グループの人数

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
99	104	20 (8)	20 (1)	19 (0)	24 (2)	21 (0)	34 (9)	35 (1)	35 (1)	11	93	11
00	89	17 (8)	16 (2)	20 (3)	18 (2)	18 (0)	26 (10)	29 (3)	34 (2)	15	72	14

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.

者はいない。ところが、99年にはHMに2人、00年にはHMに2人、Mに3人の不合格者がいる。国試を占う総合卒試としては、不合格者になるべく高い比率でLに属することが、全体として良い試験と言える。総合卒試が単なる国試の模試であった98年以前と、これによって卒業が決まる99年以降では、学生の対応の仕方が異なるのであろうか。

各問について、正答を1、誤答を0として、総点との間の相関係数 ( $r$ )、「母集団正答率の間に差がない。」という仮説の下に、Tの正答率とLの正答率の差を標準化した値 ( $u$ )、Sの正答率とFの正答率の差を標準化した値 ( $z$ ) を求める。不合格者数があまり多くないため、正答率が0.5から離れると、2項分布の正規分布による近似があまり良いとは言えないが、1つの指標として、 $z$ の大きい問が国試の合格者と不合格者を識別する性能が高いと考える。しかし、 $z$ は国試の合格者が分からないと

求まらない値であるから、合否予測には  $r$  や  $u$  を使い、総点の高い者が実力があるという前提に、 $r$ 、 $u$  についても値が大きい程識別性能が優れていると考える<sup>5)~7)</sup>。

また、統計学の知識のない者でも理解出来る簡便なものとして、L~Hの5群を  $i = 1, \dots, 5$  で表し、 $i$  群の正答率を  $x_i$  とし、 $i < j$  かつ  $x_i > x_j$  が成り立っているとき、すなわち総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  $i, j$  ( $i < j$ ) の組合せに対する転位の数 (転位数と呼び、 $tr$  で表す。) を導入する。もちろん、転位数が小さい程識別性能が優れているとする。さらに、 $a, b, c$  の3群を正答率の高い順に並べたものを、どの辺りで転位が起きているかを示す指標として用いる。5群では120通りの並べ方があるが、3群なら6通りであるから分かりやすい。なお、直接識別性能を示すものではないが、正答率を  $t$  (%) で表す。

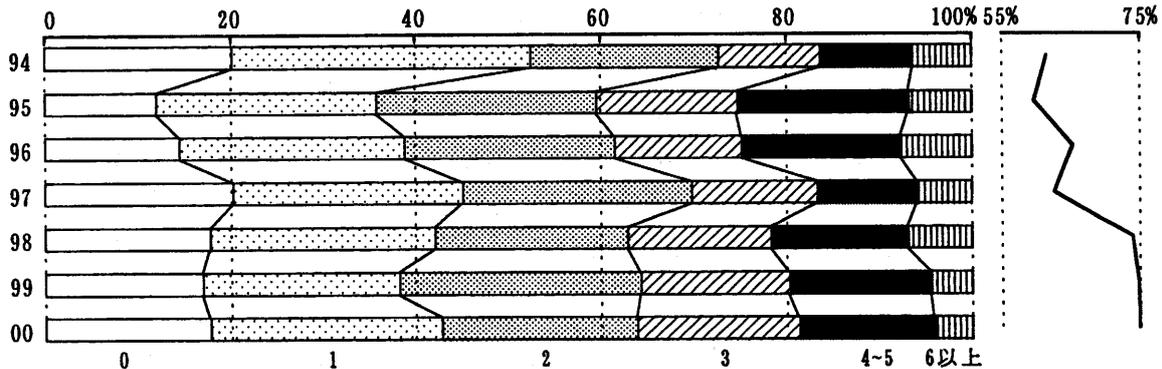


図1 転位数別の問の比率 (正答率100%及び0%の問を除く)

平均正答率

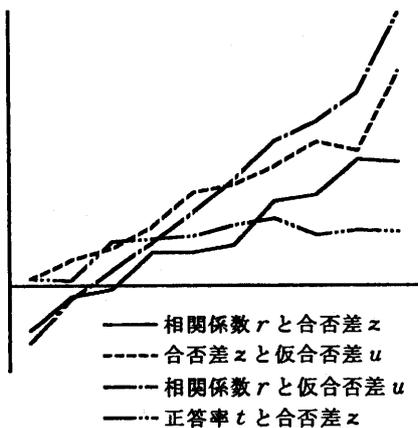


図2-1 合否差 ( $z$ )、仮合否差 ( $u$ )、相関係数 ( $r$ )、正答率 ( $t$ ) の関係 (99)

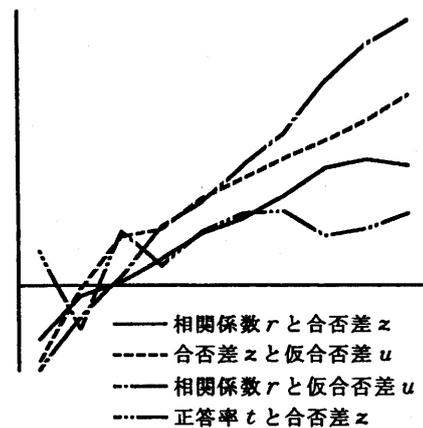


図2-2 合否差 ( $z$ )、仮合否差 ( $u$ )、相関係数 ( $r$ )、正答率 ( $t$ ) の関係 (00)

さて、図1を見ると、転位数の小さい間の比率が最も高いのは94年で、2番目は97年である。00年は転位数が1以下の間の比率が3番目で、4以上の間の比率は3番目に低い、6以上の比率は最も低い。また、99年も6以上が00年に次いで低い。全体として極端な間が少なくなってきたことを示しているとも考えられるが、94、97、00年に共通しているのは、国試の不合格者が多いということで、出来の良い者とそうでない者の差がはっきりしているため、転位が起きにくい状況にあるとも考えられる。ちなみに、正答率の全問平均を見ると、97年までは60%前後であるが、98年以降は毎年ほぼ75%である。模試的なものであった総合卒試が直接卒業判定に関係するようになったので、出題者が意識的に易しくしたことや学生がより勉強するようになったことが考えられるが、99、00年の国試が振るわなかったところを見ると、その勉強法が卒業を強く意識したもので、丸暗記的な勉強に終始した者も少なからずいるのではなかろうか。

図2は  $r, z, t$  を10階級に分け、各級の中央の値を横座標とし、 $r, z, t$  それぞれの各階級に属する問の  $z$  又は  $u$  の値の平均値を縦座標として点を取り、折れ線で結んだものである。本質的に同じものを異なる指標で表した  $r$  と  $u$  のグラフがほぼ直線であるが、 $z$  と  $u$  及び  $r$  と  $z$  も比較的直線に近く、可否予測に  $r$  及び  $u$  の大きな間を用いることが有効であることが分かる。また、 $t$  と  $z$  の関係を見ると、99、00年のどちらも、 $t = 70$  辺りに  $z$  のピークがあり、 $t \leq 20$  を満たす間は  $z$  が極端に小さく、負

となる間も少なからずあることが分かる。

ちなみに表4を見ると、 $z$  と  $r$  の間の相関係数は99年が0.533、00年が0.522、 $z$  と  $u$  の間は99年が0.569、00年が0.532で、いずれもかなりの相関がある。しかし、その推移を見ると、97年をピークに99、00年は下降傾向にあり、特に  $z$  と  $r$  の間は00年が7年間で最も低く、次いで99年である。この両年は全体的に正答率が高く、はっきりした差がつきにくいことも原因の1つとして考えられるが、L以外に属している不合格者が比較的多いことも関係がある。  $z, r, u$  と  $t$  の間は、年によってほとんど相関が見られないこともあるが、大体低い相関である。この外、 $z$  と  $tr$  及び  $z$  と  $abc$  の関係等も示したが、 $z$  及び  $t$  以外の4つは本質的に同じものを測っているわけであるから、それら4つの間の関連係数は大きい。

99年の  $z$  及び00年の  $r, z, u$  のベスト3及びワースト3の正答率曲線を図3に示した。98年以前のものについては文献4)、6)を見ていただきたい。  $r$  のベスト3の間はいずれも  $tr = 0, z \geq 2.33, u \geq 2.93$  で、99年 ( $z \leq 1.99$ ) より良くなっている。これに対して、 $r$  のワースト3は99、00年とも、 $r$  はもちろん  $z, u$  ともにすべて負で、 $tr = 6 \sim 8$  となっている。00年の②(ワーストから2番目、以下同様)は  $t = 86.5$  で、Lの正答率が100%の外、LMからHまでほとんど差がなく、 $z$  を見ると負の識別性能もないことが分かる。①はLからHM ( $t = 0$ ) まで単調に減少し、③はL、LM ( $L < LM$ ) を除いていずれも  $t = 0$  である。①の正答率13.5%は選択肢の中で

表4 各指標間の関連係数等

	94 (n=326)		95 (n=321)		96 (n=322)		97 (n=333)		98 (n=313)		99 (n=313)		00 (n=311)	
	$\chi^2$	c.c.												
$t-r$		0.247		0.179		0.258		0.267		0.296		0.206		0.143
$z-t$		0.314		0.180		0.259		0.380		0.308		0.169		0.083
$t-u$		0.258		0.267		0.320		0.411		0.217		0.226		0.204
$z-r$		0.547		0.591		0.625		0.775		0.619		0.533		0.522
$r-u$		0.785		0.819		0.824		0.869		0.838		0.899		0.823
$z-u$		0.533		0.547		0.599		0.853		0.400		0.569		0.532
$z-r$	87.2	0.366	77.9	0.348	126.0	0.442	145.0	0.467	100.2	0.400	71.2	0.337	56.6	0.302
$z-u$	79.6	0.349	74.4	0.340	95.0	0.384	187.8	0.531	40.8	0.255	75.3	0.347	44.6	0.268
$z-tr$	41.1	0.251	54.1	0.290	53.2	0.288	89.1	0.366	22.2	0.188	29.5	0.217	17.5	0.168
$z-abc$	46.6	0.267	60.9	0.308	41.6	0.254	74.2	0.334	20.0	0.179	29.0	0.215	35.0	0.237
$r-tr$	213.7	0.573	205.6	0.566	155.3	0.491	210.0	0.562	151.0	0.491	166.9	0.516	136.3	0.468
$r-abc$	218.2	0.579	217.7	0.583	160.9	0.500	184.6	0.526	151.5	0.492	134.3	0.463	123.0	0.445
$u-tr$	130.5	0.447	136.1	0.460	108.2	0.410	168.0	0.502	138.5	0.470	130.3	0.456	88.4	0.377
$u-abc$	132.7	0.451	122.7	0.437	95.3	0.385	129.0	0.440	115.1	0.429	79.6	0.357	66.1	0.326

c.c.: 上から6行目までは相関係数、7行目以下はクレマーの関連係数。  $\chi^2$  の自由度は4。正答率100%及び0%の間を除く。

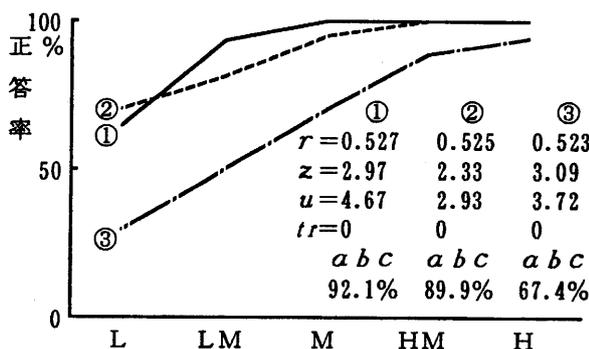


図3-1  $r$  のベスト3 (00)

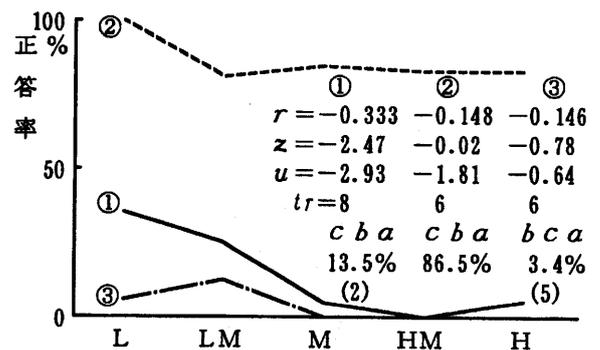


図3-2  $r$  のワースト3 (00)

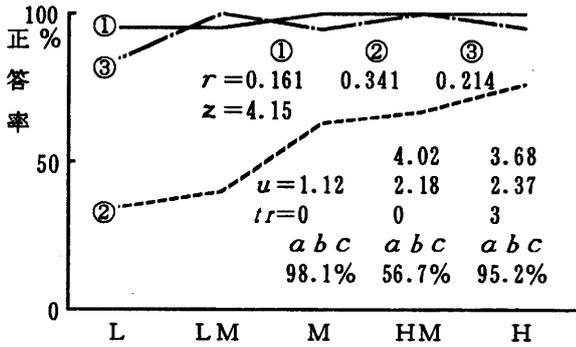


図3-3 zのベスト3 (99)

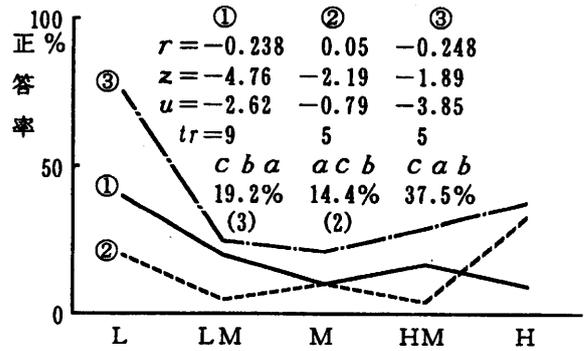


図3-4 zのワースト3 (99)

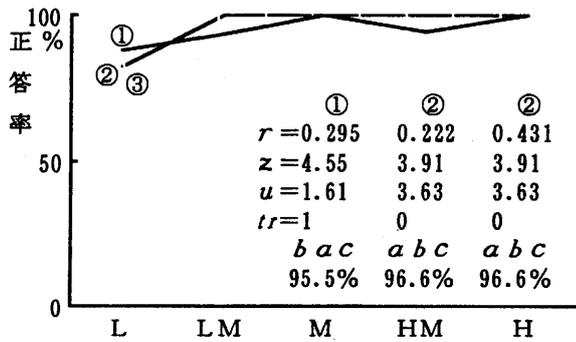


図3-5 zのベスト3 (00)

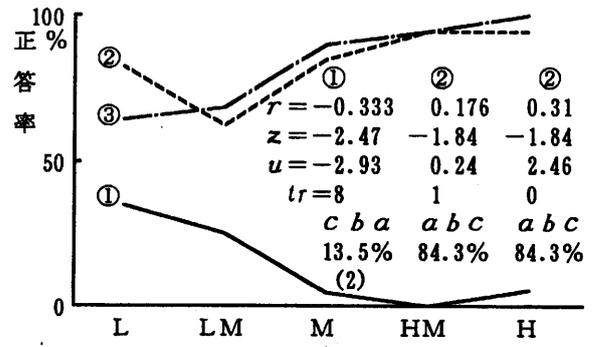


図3-6 zのワースト3 (00)

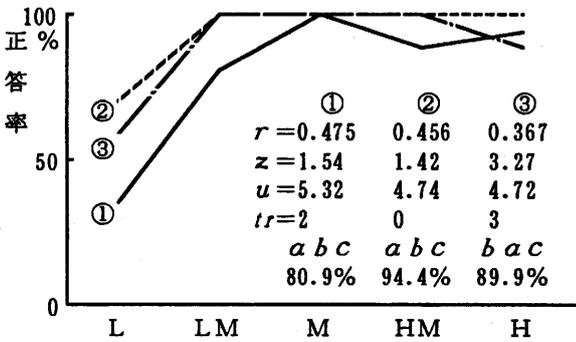


図3-7 uのベスト3 (00)

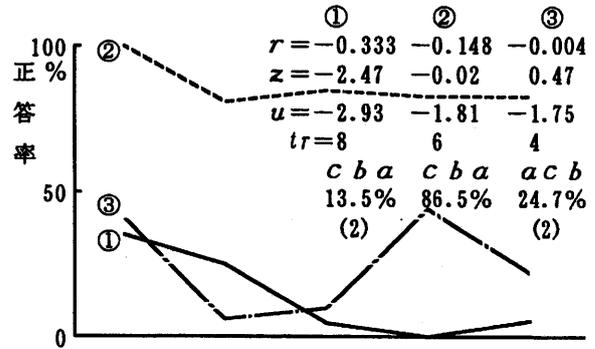


図3-8 uのワースト3 (00)

2番目に大きいのが、③の正答率3.4%は最小である。  
 zのベスト3は、99年の② ( $t=56.7$ )を除いて、いずれも  $t > 95$  である。正答率が0.5から離れるとzが大きくなる傾向があるが、00年の①は誤答者4人全員が不合格、②及び③は誤答者3人全員が不合格でLに属している。すなわち、「これらの問が出来ない人は必ず落ちる。」という点では確かに識別性能が高いと考えられるが、これらが出来た人も沢山不合格になっている。99年の①は誤答者2人が2人とも不合格でかつLに属している。②は不合格者11人全員が誤答で、合格者は 正答：誤答 = 59：34 である。また、正答率が0.5に近く、rもuも①よりはるかに大きく、常識的には識別性能が非常に高い問と言える。ちなみに、 $30 \leq t \leq 70$  を満たす問に限ると、00年のzのベスト3の正答率曲線等は図4のようになるが、この条件を満たす問はzの全体のベスト10の中に1問も入っていない。

zのワースト3を見ると、99年はいずれも正答率が低

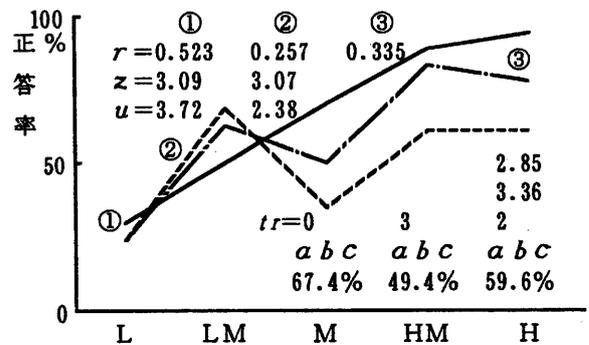


図4 正答率30%~70%のzのベスト3 (00)

く、①は  $t_r = 9$ 、②、③は  $t_r = 5$  で、いずれも負の識別性能と考えられる。00年の①は正答者12人のうち5人が不合格者で、Hに1人、HMに0人、Mに1人、LMに4人、Lに6人となっている。すなわち、総点の低い者程この間に正答する確率が高い傾向が見られ、転位数は

8である。②はどちらも誤答者14人全員が合格しているが、LあるいはLMに誤答者が集中し、 $0 < r$ ,  $0 < u$ ,  $tr = 0.1$ で、何とも理解し難い問である。

$u$ のベスト3は、当然のことながら、いずれもLの正答率が低く、LM~Hでは高い正答率で、その間にあまり差がなく、 $r$ も比較的大きいものばかりである。また、 $u$ のワースト①は $z$ のワースト①と同じ問であるが、②、③は $r$ ,  $z$ の絶対値がいずれも小さく、負の識別性能があるとは言い難い。

3 最も多く選ばれた選択肢が正解と異なる問について

「各問の正解が、5つの選択肢の中で何番目に高い比率で選ばれているか。」を表5に示した。99年は正解が最

表5 正解選択者比率の順位別問数

	1位		2位		3位		4位		5位		計
	99	00	99	00	99	00	99	00	99	00	
A	83	81	3	4	1	0	0	1	0	1	87
B	79	79	6	6	2	1	0	2	1	0	88
C	18	17	2	2	1	2	0	0	0	0	21
D	29	26	1	1	0	2	0	0	0	0	30 <sup>1)</sup>
E	42	42	6	8	3	0	0	0	0	0	51 <sup>2)</sup>
F	41	38	5	8	1	2	1	0	0	1	49 <sup>3)</sup>
計	292	283	23	29	8	7	1	3	1	2	326
%	89.6	86.8	7.1	8.9	2.5	2.1	0.3	0.9	0.3	0.6	

1), 2): 00年にすべて正解の間1, 3): 99年にすべて正解の間1を含む。

も多くの受験者から選ばれた問が約90%を占め、00年もその比率が86%を越えている。両年とも、残りの問の約70%で正解が2番目となっているが、00年は4番目と5番目の問を合せると5問になる。

00年に1位でない問は41問であるが、これらの問について、最も多く選ばれた選択肢を正解としたときの $r$ ,  $u$ ,  $z$ の符号の変化を表6に示した。 $r$ 及び $u$ は+のまま

表6 選択者比率最大の選択肢を正解としたときの符号の変化

	-	+	+	+	-	-	+	-
$r$		7		15		0		19
$u$		9		14		0		18
$z$		10		3		3		25

-+: -から+に変わることを示す。他も同様。

ま不変なものが3分の1強であるが、-のまま不変なものは皆無である。 $z$ は41問中25問で+から-へ、10問で-から+へ変化している。-から+に変わった問は「最も多くの受験者から選ばれた選択肢こそ正解ではないか。」と疑われても仕方がないが、この10問中 $r$ ,  $u$ ともに-から+に変わったものは2問だけで、残りの8問は変化前も変化後も $|z| \leq 1.10$ を満たしているから、出題者の主張を覆す程の統計的根拠はない。

$r$ ,  $u$ ,  $z$ のすべてが-から+に変わったうちの1問は、 $z$ のワースト①の問で、もし選択された比率の最も高い選択肢を正解とするならば、正答率曲線は図5の通りで、 $z = 2.83$ ,  $r = 0.392$ ,  $u = 3.40$ である。すなわち、識別性能がかなり高い問となる。もし、正解が間違っていないのなら、「教育上問題があった。」と言わざるを得ないが、

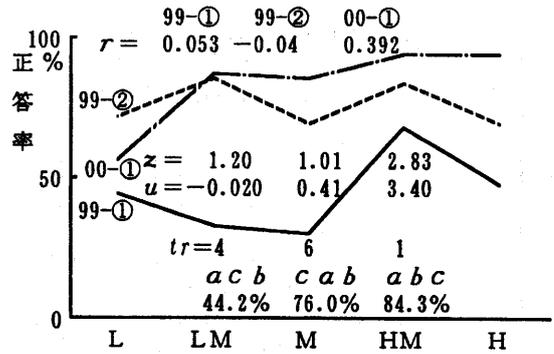


図5 zのワーストの最大比率選択肢を正解と仮定した場合

出題者の出張等により正解を確認出来ない問が毎年1~2はある。もう1問は、 $z$ が $-1.54 \rightarrow 1.16$ ,  $r$ が $-0.116 \rightarrow 0.246$ ,  $u$ が $-0.52 \rightarrow 1.52$ と変る。また、99年の $z$ のワースト①は、正解とされる選択肢が選ばれた比率が3番目、②は2番目である。しかし、①と②について、選択された比率が最高の選択肢を正解としたとき、 $z > 1$ となるが、 $r$ 及び $u$ の絶対値はともに小さく、 $tr = 4, 6$ で、決して識別性能が高いとは言えない。

4 総合卒業試験による国試の合否予測と禁忌肢

総合卒試で下から10番目までの者の約半数が不合格となっているので、毎年下から10人を卒延にすれば、国試の合格率は大幅に上がる。しかし、合格の確率が2分の1もある者を簡単に卒延にするわけにはいかない。そこで、 $i$ の値を変化させ、 $0.02(i-1) < r$ という条件を満たす問だけによる得点を用い、「母集団平均に差がない。」という仮説の下に、その得点分布が正規分布であるとして、合格者の平均と不合格者の平均の差を標準化した値( $x$ )の変化を考察する。もし、総点を用いたときより、制限を加えて得られた点を用いた方が例年 $x$ の値が大きければ、新しい年のその点による下位の何人かを「国試不合格の可能性が大きい者」と予測する。

不合格者数に大きな違いがあるから、年度間を比較するのは適当ではないが、図6に示したように、94~96年は $i = 15 \sim 16$  すなわち  $0.28 \sim 0.30 < r$  通りに $x$ の

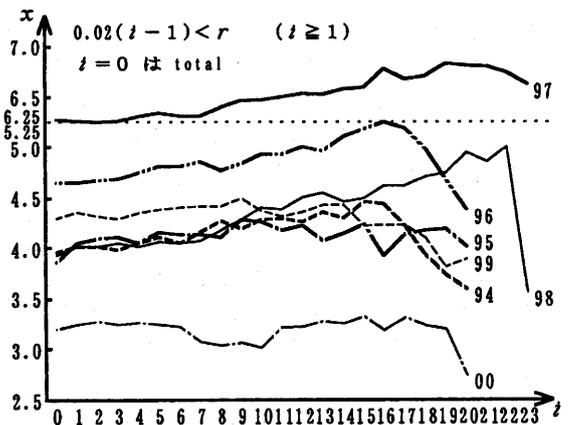


図6 rが条件を満たす問の得点の合格者平均と不合格者平均の差を標準化した値xの条件に伴う変化

ピークが見られ、97、98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97年は58、98年は45、以下同様)まで増大している。99年は  $x$  がほとんど変化せず、 $i = 10$  にピークがあるが、 $i = 14, 15 (105, 89)$  のとき2番目に大きい。また、00年もあまり変化は見られないが、ピークは  $i = 16 (99)$  のときである。7年間を通して、 $i = 16$  辺りに  $x$  のピークがあると考え、 $0.3 < r$  を満たす問による得点で順位を付け、下から10位までに入る不合格者の数を総点によるものと比べると、94年(不合格者17人)は5人(総点)→7人( $0.3 < r$  を満たす問による得点)、95年(12人)は4人→6人、96年(7人)は5人→6人とその数を増やし、97年(16人)は最初から9人、98年は不合格者が2人だけで、99年(11人)は5人で変わらず、00年(15人)は4人→5人となっている。

ところで、HM、Mに属しながら不合格になった原因として禁忌肢の選択が考えられるが、表7を見ると、

表7 グループ別に見た禁忌肢選択状況

	L	LM	M	HM	H	計
人数	17 (8)	16 (2)	20 (3)	18 (2)	18 (0)	89 (15)
0問	3 (2)	4 (1)	2	5 (1)	6	20 (4)
1問	10 (4)	6 (1)	14 (2)	5 (1)	7	42 (8)
2問	4 (2)	5	4 (1)	7	5	25 (3)
3問	0	1	0	1	0	2
合計	18 (8)	19 (1)	22 (4)	22 (1)	17	98 (14)
平均	1.06 (1.00)	1.19 (0.05)	1.10 (1.33)	1.22 (0.05)	0.94	1.10 (0.93)

合計：延べ問数。( )内は不合格者。

合卒試受験者全員で1人平均1.10個の禁忌肢を選んでいなのに対し、不合格者の平均は0.93個である。また、L～Hのグループ別に見て、下位グループは誤答率の高い者の集団であるから、その平均個数がいくらか多くなっても当然であるが、下位グループの者が多く選ぶという傾向は全く見られず、禁忌肢選択状況は合否予測に役立たない。なお、00年は禁忌肢を含む問が7つあり、そのうちの2問は禁忌肢が2つある。

$0.18 (i - 1) < u$  を用いても、 $r$  の場合と同様に、合

表8 総合卒試成績及び国試合否別に見た高校調査書及び学内成績等

	総合卒試順位	現12学 役浪浪士	調査書 ④ABC	評定平均 値の平均	学内成績順位の平均					
					1年	2年	3年	4年	5年	6年
a 上位10位までの者(10人)	1～10	4312	5140	4.35	22.3	10.5	9.4	9.7	14.1	18.3
b 上・中位の不合格者(5人)	19～47	1112	2102	4.14	37.8	58.6	62.4	55.0	50.2	60.6
c bの者と最も近い合格者(5人)	19～49	2210	3110	4.56	51.2	45.0	47.6	54.4	51.0	55.8
d 下位の不合格者(5人)	64～78	1130	0032	3.62	56.0	55.2	61.0	66.6	55.0	70.8
e dの者と最も近い合格者(5人)	65～76	2030	1031	3.82	58.6	65.2	55.0	55.0	66.8	62.0
b + d (10人)	19～78	2242	2134	3.88	46.9	56.9	61.7	60.8	52.6	65.7
c + e (10人)	19～76	4240	4141	4.19	54.9	55.1	51.3	54.7	58.9	58.9
f 下から10位までの不合格者(4人)	82～87	0301	1120	4.10	71.6	67.3	53.5	71.0	62.3	66.0
g 下から10位までの合格者(6人)	79～88	1302	0231	4.02	79.2	79.2	77.2	75.8	52.5	60.7
b + d + f (14人)	19～87	2543	3254	3.94	52.5	59.9	59.4	63.7	55.2	65.8
c + e + g (16人)	19～88	5542	4372	4.13	62.4	64.1	61.0	62.6	56.5	59.6

学年成績順位：1年次は一般教育科目の単位数をウェイトとした平均点による順位。2年次及び3年次は基礎医学系科目、4年次及び5年次は臨床医学系科目、6年次は臨床検査医学、臨床講義、CPC、総合医学概論の4科目のウェイトを同じにした平均点による順位。2浪は2浪以上の者、学士は88年以前に高校を卒業した者。dは下から10位までの者を除く。総合卒試及び6年次の順位は卒延となった者を除く。

否予測における上記の確率を10%程度高めることが出来るが、問全体の改善がなければ、この程度が限界である。もちろん、総合卒試の改善は、単に合否予測力を高めることだけを目的とするのではなく、学生の学力の向上を第一義に考えるべきものである。

### 5 国試不合格者の6年間の成績等

00年の国試不合格者を、HM又はMに属する者(b,5人)、LM又はLに属するが卒延になった89位の者を除いて下から10位までには入っていない者(d,5人)及び下から10位までに入っている者(f,4人)に分け、b及びdの者と比較するため、同順位又は最も順位の近い合格者を選んでグループを作り(c及びeとする)、fに対しては下から10位までに入っている合格者(6人)をgとし、参考までに上位10位までの者をaグループとする。a～gのグループごとに、現役、1浪、2浪以上、学士(88年以前に高校を卒業した者)の人数、高校調査書の学習成績概評の④、A、B、Cの人数及び評定平均値の平均を求め、また1年次から6年次(総合卒試と学生によって評価者が異なる選択実習を除く。)までの成績の平均点による順位の平均を求めて、それらを表8に示した。個人の学年平均点は、1年次が一般教育科目の単位数をウェイトとした加重平均、2年次及び3年次が基礎医学系科目、4年次及び5年次が臨床医学系科目、6年次が臨床検査医学、臨床講義、CPC、総合医学概論の4科目の各科目のウェイトを同じにした平均である。なお、順位でグループ分けしているため、回帰効果が見られるのは当然である。

1期生(80年入学)から10期生(89年入学)までの9期生を除く約900人についての調査結果によると、ストレートには医師になれなかった者の割合が、④では24.3人に1人、Aでは10.1人に1人、Bでは7.1人に1人、Cでは3.6人に1人である<sup>3)</sup>。また、平均して現役は浪人より入学後の成績が良く、浪人年数の長い者程良くないことが分かっている<sup>4)</sup>。表8では、aは現役の比率が高く、

④が半数でCは皆無である。少数例なので、これだけで何らかの結論めいたものを出すわけにはいかないが、 $c + e + g$  は  $b + d + f$  より現役の比率が高く、④、Aの比率も高い。また、評定平均値でも同じ傾向が見られる。ちなみに、④の者は19人中3人、Aは14人中2人、Bは42人中5人、Cは12人中4人が不合格で、文献3)の結果と比べると、④、Aでその比率が高くなっている。なお、国立で推薦を実施する大学が増えたためか、入学者の中での④の占める割合は、その当時に平均 8.1%であったものが、00年には 21.6%と大幅に上昇している。

学内成績順位を見ると、aは1年次から抜群に良く、3年次までは、 $b + d + f$  の方が  $c + e + g$  より良いが、その差は次第に小さくなり、4年次には逆転している。5年次に再逆転しているが、ほとんど差はなく、6年次には  $c, e, g$  がそれぞれ  $b, d, f$  を上回り、平均で 4.8~8.8 の差をつけている。すなわち、1年次から5年次までの成績を考察しても、総合卒試のHMあるいはMに属する者が不合格となった原因の究明には役立たないが、総合卒試以外の6年次の成績は注目に価する。

HM及びMに属する5人について、6年次の成績を中心に個別に考察する。1人の学士は1年次から5年次まで単調に順位を下げ、CPCが60点である。ちなみに、CPCが70点未満の者はわずかに5人である。もう1人の学士は6年間を通して成績は良いが、調査書がCで、社会人としての期間がかなり長い。2浪で調査書がCの1人は臨床講義が60点である。この科目で60点の者は13人いるが、そのうちの5人が不合格者である。調査書が④の2人はともに臨床検査医学が60点で、そのうちの1人は臨床講義も60点でかつ1年次から5年次までの最高の順位が85位である。もう1人も6年次の順位では下から4分の1に入っている。なお、臨床検査医学では60点の者13人中4人が不合格、70点未満の者20人中7人が不合格である。このように、社会人としての期間の長い1人を除く4人は、総合卒試で実力を上回る結果が出たようである。ともあれ、国試も入試と同様に一発勝負であるから、当日の体調等を含め、常に運に左右され、不合格の原因を完全に解明することは不可能である。

## 謝 辞

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り、本論文の原稿に目を通されて貴重なご意見を下さった、前恵泉女学園学園長の川田殖先生、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下さい、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、日ごろのご支援と合せて、感謝の意を表したい。

## 文 献

1) 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, 4:6~13.

- 2) 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可否の関係. 大学入試研究ジャーナル, 5:39~49.
- 3) 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可否の関係. 大学入試研究ジャーナル, 6:76~83.
- 4) 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, 14:50~60.
- 5) 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.
- 6) 平野光昭 (1998) 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. 山梨医科大学紀要, 15:97~104.
- 7) 平野光昭 (1999) 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. 大学入試データの解析 (柳井晴夫, 前川真一編). 現代数学社, 第IV部これからの入試の1, 216~229.

## Abstract

### **The Correlation between the Results of the Examinations in Yamanashi Medical University (YMU) and the Success or Failure in National Examination for Medical License (NE) – Prediction of the Pass-rate of NE with the Special Aid of the Comprehensive Graduation Test (CGT) –**

**Teruaki HIRANO**

The elevation of the quality of successful applicants and the improvement of the contents of education result in the increase of medical students who are to go straight on to doctors in due course of six years and the birth of promising doctors full of human excellence. However, as the pass-rate of NE is often considered as a landmark of the contents of medical education in the universities and colleges, we cannot disregard the elevation of the pass-rate of NE.

Therefore, we contrived the system of data analysis and endeavored to improve it for the past several years both to examine the discriminatory powers of questions in CGT which has close affinities with NE and, using the particularly discriminatory questions, to predict the pass-rate for any student.

In this essay, the author, applying these newly improved methods, analyzes the data of CGT for the students graduated in 1999 and 2000, and, considers the correlation between the results of the examinations in YMU and the success or failure in NE.

In case we could obtain cooperation from many universities, by storing in computers the 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.

---

Department of Mathematics