

総合卒業試験による医師国家試験の 合否予測はどこまで可能か (その2)

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、将来に向けて有望な医師を誕生させることになるが、世間では医師国家試験（以下国試と呼ぶ。）の合格率を大学教育の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業試験（以下総合卒試と呼ぶ。）の間の中から、項目反応の理論等を応用して、何種類かの基準で、識別性能の高いものを選び、これらの問だけによる得点で、合否の予測がどこまで可能か、今年（93年度）～95年度に卒業した者のデータを用いて、昨年（92年度）の方法を改善し、より一層予測力を高めることが出来た。

キーワード：総合卒業試験，医師国家試験，項目反応，識別性能，合否予測

1 はじめに

80年に1期生を迎えた本学は、96年に11期生を送り出し、この間に1100人近い医師を誕生させた。最近の医師国家試験（以下国試と呼ぶ。）の大学別合格率による順位を見ると、92、93年は極めて上位であったが、一昨年及び昨年は低迷し、今年は再び上位となった。1人でも多く国試に合格させることだけが大学の教育目標でないことは言うまでもないが、世間ではこの合格率を医学教育の評価基準の1つとしており、医学教育学会にも国試に関する専門委員会が設けられているなど、多くの会員が国試の合格率に強い関心を持っている。学内には2年連続の不振を憂慮している向きが多かったが、今年は向上への対策が功を奏して一安心しているところであろう。しかし、表1に示したように、本学は6年間で卒業する者の比率が高く、6年間でストレートに医師となった者は、開校以来11年連続して入学者の80%以上で、全国で

も極めて高い水準にある。

入学する学生の質を高めること及び教育の質を向上させることは、単に6年間でストレートに卒業する者の数を増やし、国試の合格率を高めるだけではなく、将来に向けて有望な医師を誕生させることにもなり、ともに重要なことである。入学する学生の質を高める要因については、これまでに数多くの研究成果を発表してきた^{1)~8),11)}。そして、推薦選抜を94年から実施し、追跡調査では「一般選抜による入学者より学内成績が良い。」という結果を得ている^{12),14)}。

ところで、我々が国試の成績に強い関心を持つようになった動機はこれだけではない。本学は開校以来絶えず入学者選抜方法の改善に努め、他大学の改革の影響も常に受けているから、入学した学生の質も年々変化しているはずである。しかし、医学部では大部分の授業が学年単位で行われているので、学内成績は同一学年での相対評価となる傾向が強く、入試改革の成果を学内成績で見るとは難しい。そこで、全国値と比較できるものとし

表1 倍率・入学時の学力レベル・卒業生数・国試合格率等の年度間の比較

入学年度	入学者数	実質倍率	z-値	卒業年	卒業生数	卒延生数 (a)	国 試				修正不合格者数 $x = a + \frac{14.5b}{100-c}$
							合格者数	不合格者数 (b)	合格率 (d%)	全国合格率 (c%)	
80	100*	3.6	1.46	86	90	10	84	6	93.3	86.6	16.5
81	100	2.3	1.50	87	97	3	94	3	96.9	86.2	6.2
82	100	3.1	1.70	88	96	4	88	8	91.7	81.2	10.2
83	100*	2.6	1.47	89	91	9	80	11	87.9	88.0	22.3
84	100	2.9	1.61	90	97	3	86	11	88.7	82.9	12.3
85	100*	1.9	1.28	91	93	7	85	8	91.4	84.3	14.4
86	100	3.8	1.47	92	90	10(8)	86	4	95.6	84.0	11.6**
87	100	9.6	1.66	93	92	8	90	2	97.8	90.1	10.9
88	100	6.0	1.63	94	93	7	80	13	86.0	86.2	20.7
89	100	7.1	1.64	95	93	7	82	11	88.2	86.0	18.4
90	100	6.8	1.38	96	87	13	85	2	97.7	89.3	15.7

* 沖繩留学生1名を除く。 ** () 内の補正值8を使った。100-cは全国不合格率で、14.5は全国不合格率の88年までの9年間の平均である。

て、国試の大学としての成績すなわち合格率（年度によって全国合格率が違っているので、これによって補正した値、表1の x ）を用いてその成果を見ることにしたのである。入学時の学力レベルを年度間で比較するとき、個別学力検査ではなく、大学入試センター試験の全国値と比較した成績（表1の z ）を用いるのもこれと同じ理由による。

さて、医学部を卒業することを前提に行われる国試も「一発勝負」の試験であるから、教育の質の向上の結果と称して、卒業判定基準を甘くすれば、ストレート卒業者数が多くなるので、6年間でストレートに医師となる者の比率が高まるが、これに伴って合格率は一般に低下する。もし何らかの方法によって国試の合否を高い確率で予測出来れば、不合格の可能性の高い者を卒業させずにもう1年間勉強させることによって、6年間でストレートに医師となる者の比率をあまり下げずに、国試の合格率を飛躍的に高め、その順位を全国のトップに押し上げることも可能になる。また、学生は卒延になったことを納得し、反省もするだろう。

これまでの追跡調査の結果から、国試の模試とも言われている総合卒業試験（以下総合卒試と呼ぶ。）の成績は、卒業試験全体の平均値より、国試の合否との関連が強いことが分かっていた（総合卒試を除く各科目の卒業試験の結果によって卒延となる者が出るので、卒業試験全体の平均値には「選抜効果」が働いている。）ので^{2),3),5),6)}、総合卒試によってこの予測が出来ないか考えた。

国試不合格者の多くはこの総合卒試の成績が下から10番以内であるが、年によっては全体の中以上の者もあり、その原因の解明に苦慮している。一方、正答率が10%を割る問が毎年何問か見られ、これらの問は学生の実力を判別する性能が欠けているのではないかと疑問を感じていた。そこで、総合卒試の問の中から、項目反応の理論等を応用して、何種類かの基準で、識別性能の高いものを選び、これらの問だけによる得点によって、この予測がどこまで可能か追究してみることにし、昨年は94年度卒の者について考察した⁹⁾。その後、平成7年度科学研究費補助金・総合研究（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」の研究集会で多くの方からコメントをいただいたので¹⁰⁾、今年は昨年の方法を改善し、93年～95年度に卒業した者について考察する。

2 総合卒試の各問の実力識別性能

総合卒試の327問の中に正答率が10%を割るものが毎年何問かあることは、前述の通りであるが、試験の終了直後に、受験した学生から「正解が見当たらない。」とか、「正解が間違っているのではないか。」などの問合せがあつて、正解が訂正されたり、その問が採点の対象から外されたりすることがあるようである。本試験は多肢選択方式（5つの中から）であるから、正答率が20%より低い問は一応正解を疑ってみる必要があり、特に10%を切るようなものは正解が違っている可能性が大きく、時には誤植のこともあろう。

そこで、各問（項目）が学生の実力の違いをどのように識別しているか、その実力識別性能を測るため、高野文彦氏が大学入試研究ジャーナル第2号に発表し¹⁵⁾、同氏からその応用を勧められていた項目反応の理論を用いてみることにした。まず、卒延となった者を含め、受験者を3通りの方法で、表2にあるように、それぞれグループ分けする。次に項目ごとに各群の正答率を求め、各グループ分けの下の方から上の群に向けて順に、その正答率を x_1, x_2, \dots, x_n とする。高野氏は論文の中で、「実際

$$0 \leq x_1 \leq x_2 \leq \dots \leq x_n \leq 1$$

を満たさない項目もあるが、そのような項目は「異常なものとして、分析の対象から除く。」としているが、高野氏が研究の対象とされた大学入試センター試験の場合と違い、本試験で5群に分けた場合、正答率が100%の問を除くと、この不等式をすべて満たすものは、93年が326問中65問（20%）、94年が321問中38問（12%）、95年が322問中46問（14%）である。従って、これを満たさない問を対象から除くわけにはいかない。センター試験と本試験でこのように大きな相違が見られるのは、前

表3 転位数別の問の数

年度	0	1	2	3	4	5	6	7	8	9	計
93	65	105	66	36	25	8	12	5	3	1	326
94	38	76	76	49	36	24	14	6	2	0	321
95	46	78	73	44	40	16	15	6	3	1	322
計	149	259	215	129	101	48	41	17	8	2	969

表2 各グループの年度別人数

年度	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）	
		L	LM	M	HM	H	C	B	A	F	S
93	100	20(9)	20(5)	20(0)	20(2)	20(1)	32(13)	34(2)	34(2)	17	83
94	96	19(5)	19(5)	20(1)	19(0)	19(0)	31(10)	33(2)	32(0)	12	84
95	96	20(6)	18(1)	20(0)	17(0)	21(0)	32(7)	31(0)	33(0)	7	89

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，（ ）内は不合格者数

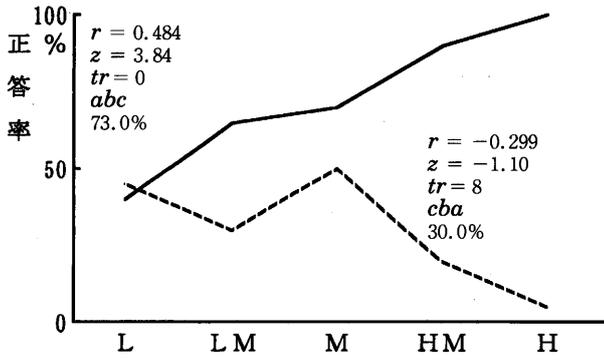


図1-1 rが最大及び最小 (93)

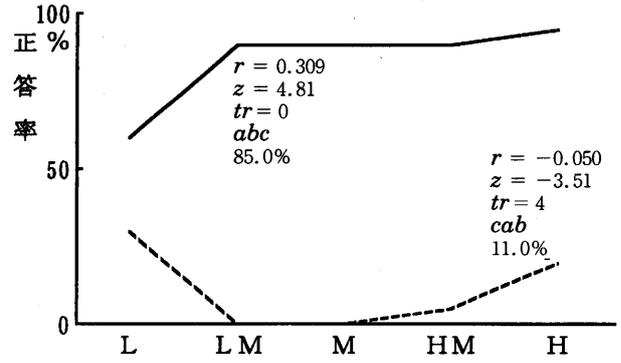


図1-2 zが最大及び最小 (93)

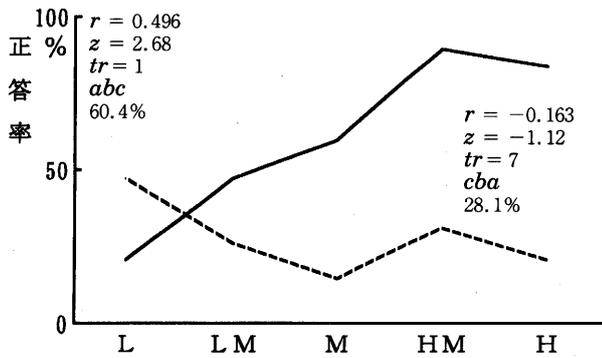


図1-3 rが最大及び最小 (94)

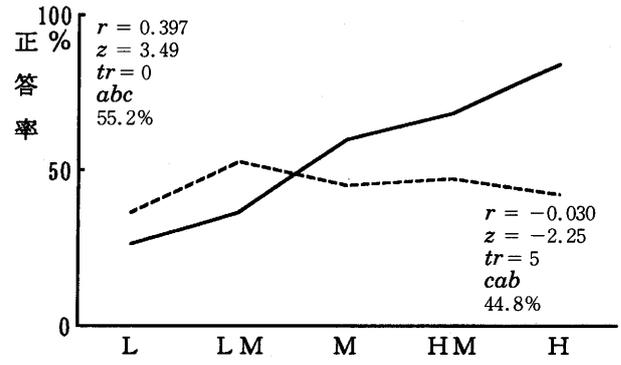


図1-4 zが最大及び最小 (94)

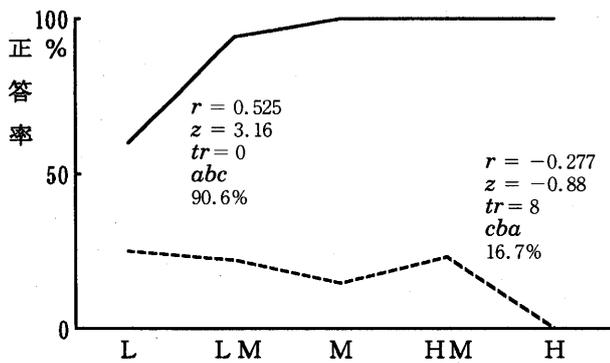


図1-5 rが最大及び最小 (95)

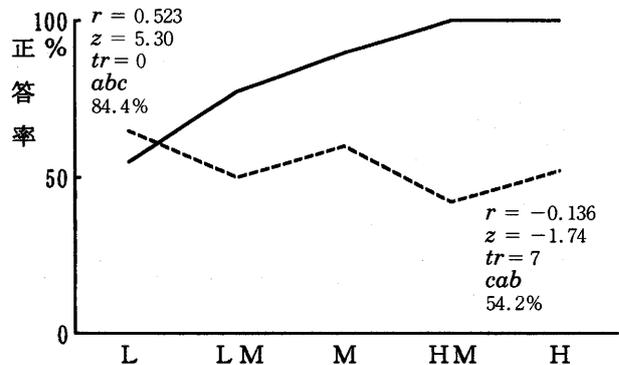


図1-6 zが最大及び最小 (95)

表4 各指標間の関係係数

	93 (n=326)		94 (n=321)		95 (n=322)	
	χ^2	c. c.	χ^2	c. c.	χ^2	c. c.
t-r		(0.247)		(0.179)		(0.258)
t-z		(0.314)		(0.180)		(0.259)
r-z	87.2	0.366	77.9	0.348	126.0	0.442
		(0.547)		(0.591)		(0.625)
r-tr	213.7	0.573	205.6	0.566	155.3	0.491
z-tr	41.1	0.251	54.1	0.290	53.2	0.288
r-abc	218.2	0.579	217.7	0.583	160.9	0.500
z-abc	46.6	0.267	60.9	0.308	41.6	0.254
tr-abc	197.0	0.550	188.9	0.542	161.0	0.500

c. c. はクレマーの関連係数。但し、() 内の数は相関係数

者は各群の人数が何万人であるのに対し後者は20人足らずであること、前者は項目数が50前後であるのに対し後者は320以上もあるので、1つの項目の総点への寄与率が小さいことによろう。各群の人数が少ないことは、安定性を低下させるが、1つの項目の総点への寄与が少ないことは分析にとって好ましいことである。

次に、実力識別性能を測る指標を4種類導入する。第1の指標は、L, LM, M, HM, Hに対する正答率の間で「 $i < j$ かつ $x_i > x_j$ 」が成り立っていれば転位が起きているとし、すべての i, j の組合せに対する転位の数(転位数と呼ぶ)である。転位数ごとの項目数は表3の通りで、1 (26.7%) 及び2 (22.2%) の項目を合せると約半数になる。以下転位数が多くなるに従って項目数は単調に減少している。言うまでもなく、転位数の少ない項目程識別性能が高いと考える。第2の指標はC, B, Aに対する $x_1=c, x_2=b, x_3=a$ の大きさの順序で、 abc から cba まで6通りあるが、 abc に近い程識別性能が高いと考える。第3の指標は、正答を $y=1$ 、誤答を $y=0$ とし求めた総得点 (x) との間の相関係数 (r) で、 r が大きい程、「 x が大きいところで $y=1$ の比率が高く、小さいところで $y=0$ の比率が高い。」という傾向がより強くなるから、識別性能が高いと考える。第4の指標は国試の不合格者群 (n_1 人) と合格者群 (n_2 人) の正答率 (\hat{p}_1 及び \hat{p}_2) の差を標準化したもので、

$$z = \frac{\hat{p}_2 - \hat{p}_1}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

で表される。但し、 p は受験者全員の正答率である。

各年度の間の中から代表的なもの4つずつを選び、第1のグループ分けによる各群と正答率の関係を図1で見ると、 r 及び z が大きい場合は、いずれも識別性能が高いことが分かるが、それらが極めて小さい項目は全く識別力がないか負の識別力で、やはり問に問題があったようである。特に r が最小の場合は正答率も低く、転位数 (tr) が7~8となっている。また、93年の z が最小の場合は正答率が11%で、中位層が低いいため $tr=4$ であるが、超難問と推測され、最下位群に何人かの正答者がいるのは「まぐれ当り」だろう。このため z はと

りわけ小さい。年度間を比べると、不合格者数の多い93年と不合格者が最下位群に集中していた95年は最大の z が大きい。

3 識別性能を測る指標間の関連

z が大きい項目程、不合格者を識別する性能が高いと言えるが、 z は不合格者が決まらないと出ない値であるから、これは予測には使えない。そこで、他の3つの各指標と z との関連を調べ、もしその関連が強ければ、その指標による識別性能の高い項目だけを用いて、合否の予測をすることが可能になる。

図2は相関係数 r と標準化された正答率の差 z 及び正答率 t (%) の関係を示したものである。正答率が10%以下の間は5問 (94, 95年) ~ 7問 (93年) であるが、それらの間の r 及び z の平均は、93年の z が0.27の外、すべて負で、いずれも正の識別性能はほとんど見られない。 t の上昇に伴って r も z も増加し、 t を10%ごとに区切った r の平均は $50.0 < t \leq 90.0$ ではあまり変わらず、 t が70.0前後のとき r は最も大きくなるようである。そして、当然のことながら t が90を越えると r は平均的に見て下がってくる。同様な z の平均は $60.0 < t \leq 70.0$ で急に大きくなり、その後も増加し、 $90.0 < t$ でも r の場合ほどには z は下がらないという傾向が見られる。

問題が難しすぎても易しすぎても r が小さくなることは容易に理解できるが、国試の合格率が80%以上であるから、多少易しくても極端でなければ、 z は必ずしも小さくならない。ちなみに、 t, r, z の間の相関係数は表4の通りで、3年間のデータを合せて考えれば、「 r と t の間及び z と t の間には相関がない。」という仮説は、1%の有意水準でも棄却される。

r を10段階に分けて z の平均をとったグラフは直線に近く、93年は $r \leq 0.0$ を満たす33問の64%に当る21問で、94年は39問の74%に当る29問で、95年は28問の71%に当る20問で $z \leq 0.0$ である。また、93年は $0.36 < r$ を満たす33問の79%に当る26問で、94年は31問の84%に当る26問で、95年は34問の85%に当る29問で $1.0 < z$ で、両者の関連は強いことが分かる。さらに、3行3列のマスを作り、 χ^2 ($\nu=4$) 及びクレーマーの関連係数を求

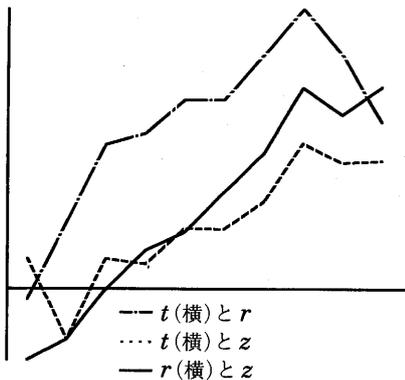


図2-1 正答率(t)と r, t と z と r と z の関係 (93)

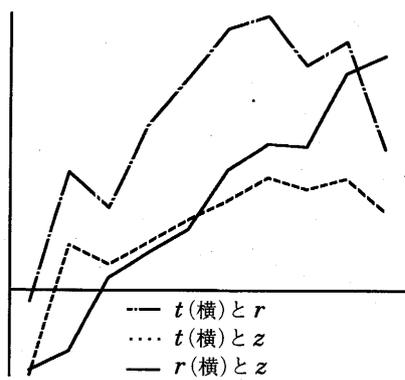


図2-2 正答率(t)と r, t と z と r と z の関係 (94)

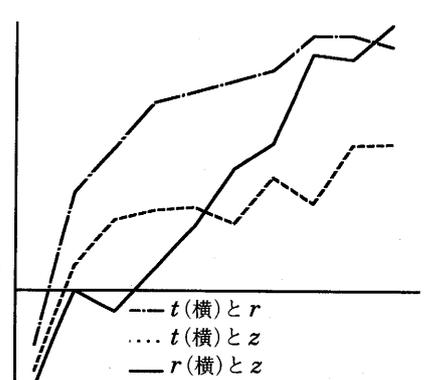


図2-2 正答率(t)と r, t と z と r と z の関係 (95)

めると表4のようになる。3行3列の分割は、それぞれなるべく等分に近くなるようにしてあるが、*abc*については、「*abc*」と「*acb, bac*」と「残り」とに分けるしかないで等分には程遠い。なお、 $v=4$ で独立性の検定をするとき、確率に関する関係式

$$P(\chi^2 > 18.47) = 0.001$$

が成り立つから、どの2つの変数の間も「独立である。」という仮説は高い水準で棄却される。また、 r と z の相関係数を見ると、いずれもクレマーの関連係数より大分大きいから、転位数あるいは*abc*の順序も z との関連は、 r には及ばないが、かなり強いと言える。そして、予測には r を用いるのが最も合理的であるが、 r だけでは視覚に訴える力が弱いから、個々の間の識別性能を見るには転位数及び*abc*の順序も役に立つであろう。

正答率が20%に達しない問や転位数の大きな問については、やはり「正解が間違っていないか。」「誤値はないか。」「内容は適切であるか。」など吟味して見る必要がある。そして、合格率が90%前後の国試の模試としては、正答率が60%~90%になるよう出題を工夫することが望まれる。また、 z の大きな過去問については関連事項を特に注意して勉強しておくとうい。

なお、93年は転位数が0又は1の問の84%に当たる143問、94年は82%に当たる94問、95年は90%に当たる111問で $0.0 < z$ であるのに対し、転位数が6以上の問では、93年が13問(62%)、94年が17問(77%)、95年が13問(52%)で $z \leq 0.0$ である。また、3群に分けた場合は、いずれも過半数の244問(93年)、166問(94年)、187問(95年)で正答率が*abc*の順に高く、*acb* ($z \leq 0.0$ の問が93年~95年にそれぞれ44%、50%、48%)の方が*bac* (同25%、24%、20%)より、*cab* (同77%、77%、64%)の方が*bca* (同43%、67%、43%)より z の平均が小さい。すなわち、 c が上に来ると z が負になりやすいという傾向が見られる。さらに、 c が上に来た場合に5群の間の転位数が多くなる傾向が93、94年に見られる。

4 識別性能の高い問による国試の合否予測

識別性能の高い問を(1) $0.165 < r$, (2) $0.196 < r$, (3) $0.256 < r$, (4) 転位数3以内, (5) 転位数2以内, (6) 転位数1以内, (7) *abc, acb, bac*, (8) *abc*, (9) $0.5 < z$, (10) $1.0 < z$ の10種類の条件をそれぞれ満たすものとして選び、卒延者を除き、国試受験者について各制限(条件)を満たす問による10種類の点数及びその点数による順位を求め、これらの点数等について、国試不合格者群と合格者群の間に素点(total)を用いた場合より高い水準で有意な差が認められるか否か検定する。それには3つの方法が考えられるが、 χ^2 分布による独立性の検定は、93年~95年の不合格者数が順に17、11、3と少ないので、これらを上位群と下位群の2つに分けても、特に95年は「これらの各成績と国試の合否は独立である。」という仮説を立てた場合の理論値(期待値)が5を下回るマスが生じて、信頼性を著しく低下させる。ちなみに、どの点数によっても、3人全員が下位群に属して、比較ができない。

残る2つの方法は順位和検定と平均値の差の t 検定である。小さい方の群(不合格者群)に属する n_1 人の順位和の和 R を求め、この R の値が「両群の母集団分布が同一である。」という仮説の棄却域に落ちるか否かを調べる(順位和検定)ため、大きい方の群の人数 n_2 が十分大きいとき、 R の確率分布は、平均 μ 、標準偏差 σ の正規分布で近似できるので、

$$z_1 = \frac{R - \mu_R}{\sigma}$$

とする(標準化)。但し、

$$\mu_R = \frac{n_1(n_1 + n_2 + 1)}{2}$$

$$\sigma_R = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

表5 各条件によって選ばれた問による点数と国試の合否の関連 (各数字は左から93年, 94年, 95年)

制限	配点			z_1			z_2			mfR			means-meanf		
total	326	321	322	3.86	3.58	2.85	3.96	3.42	3.96	75.3	76.0	89.8	8.3	8.1	17.1
$0.165 < r$	191	163	187	3.86	3.77	2.88	4.19	3.89	4.13	75.3	77.5	90.3	13.7	15.9	26.5
$0.196 < r$	168	145	163	4.02	3.66	2.87	4.28	3.78	4.23	76.3	76.6	90.2	14.8	16.3	28.8
$0.256 < r$	102	96	103	3.99	3.70	2.86	4.32	3.74	4.43	76.1	77.0	90.0	17.6	18.7	34.8
$tr \leq 3$	272	239	241	3.68	3.82	2.81	3.99	3.86	3.83	74.1	77.9	89.3	10.3	12.2	20.7
$tr \leq 2$	236	190	197	3.71	3.95	2.86	3.94	3.97	3.82	74.3	78.9	90.0	10.9	13.6	22.2
$tr \leq 1$	170	114	124	3.79	3.50	2.87	3.96	3.55	4.11	74.9	75.4	90.2	12.4	14.0	25.8
<i>acb, bac</i>	288	275	292	3.76	3.77	2.86	4.02	3.76	3.84	74.6	77.5	90.0	10.0	10.7	18.7
<i>abc</i>	224	166	187	3.82	3.73	2.86	4.21	3.83	3.85	75.0	77.2	90.0	11.9	13.7	22.6
$0.5 < z$	185	183	203	5.40	4.86	2.92	6.01	5.62	5.21	85.1	86.0	91.0	18.0	18.7	29.0
$1.0 < z$	137	128	142	5.60	5.06	2.92	6.47	6.06	5.70	86.4	87.6	91.0	21.2	22.8	34.8

配点は制限(条件)を満たす問の数、mfRは不合格者の順位の平均、means-meanfは合格者平均点と不合格者平均点の差を配点を100として換算したもの。

また、両群の成績の平均の差 $\bar{x}_2 - \bar{x}_1$ が「母集団平均の差 $\mu_2 - \mu_1 = 0$ 」という仮説の棄却域に落ちるか否かを調べるため、標本標準偏差 s_1, s_2 を母集団標準偏差 σ_1, σ_2 の近似値として、

$$\sigma_{\bar{x}_2 - \bar{x}_1} \doteq \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$z_2 = \frac{\bar{x}_2 - \bar{x}_1}{\sigma_{\bar{x}_2 - \bar{x}_1}}$$

とする。この場合も n_1, n_2 が十分大きければ、母集団分布が正規分布であるという仮定を除いても、 z_2 の確率分布は近似的に標準正規分布に従う。いま、不合格者群の標本数は十分大きいとは言えず、多少の誤差は避けられないが、ここでは total 及び10種類の各点数間の比較が問題で、 n_1, n_2 は同一年度では同じ値であり、 $n_1 + n_2$ はどの年も100前後である。なお、これらの仮説の下に標準正規分布では、関係式

$$P(z > 2.58) = 0.005$$

$$P(z > 3.08) = 0.001$$

が成り立つ。但し、 z は z_1, z_2 の総称である。これらの制限を満たす問の数(配点)、 z_1 及び z_2 の値、不合格者の順位平均、合格者の平均点と不合格者の平均点の差を配点が100点になるように換算した値は表5の通りである。

93年は転位数で制限した場合いずれも不合格者の順位平均が上がっており、 r による場合は制限を厳しくした場合に順位平均は下がっている。しかし、 z_2 の方は $tr \leq 2$ の場合を除いて大きくなっている。94年は $tr \leq 1$ の制限による場合の z_1 を除いて、total の場合より z_1, z_2 ともに大きくなっている。 r に関する条件で問を制限した場合は、 $0.165 < r$ の場合が z_1, z_2 ともに最大で、他の年度と異なる傾向を示している。制限を厳しくすると配点を100点に換算した合格者平均と不合格者平均の差は大きくなるが、制限を満たす問の数が少なくなるので安定性が低下し、一般に標準偏差が大きくなる。転位数による場合は $tr \leq 2$ の場合が最高で、他のどの場合をも上回っている。95年は卒延者が4人あり、国試不合格者が3人だけであるため、 z_1 は他年度に比べて小さく、ほとんど変化しないが、合格者平均と不合格者平均の差が大きいため、 z_2 は他年度に比べて小さくない。中でも r による制限の場合に最も大きく、3年間を通して見て、 z と最も関連の強い r による制限の場合が z_1, z_2 ともに比較的大きい。なお、 z の大きさで制限した場合に、 z_1 及び z_2 が極めて大きくなるのは当然のことである。

さて、 r に関する制限の境界値(下限)を連続的に変化させたら、 z_1 及び z_2 はどのように変わるだろうか。この様子をグラフで表したものが図3-1及び2である。横軸の0はtotalの場合で、以下 i (横軸の目盛り) に対して $0.02(i-1) < r$ を満たす問による値である。図3-2の方が変化がはっきり読めるが、93年及び95年は

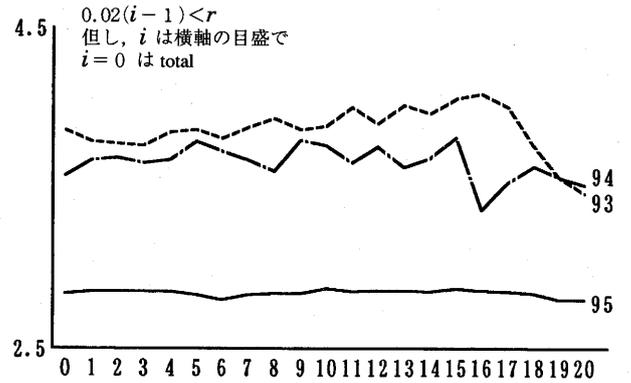


図3-1 rの大きさによって問を制限した場合の z_1 の変化

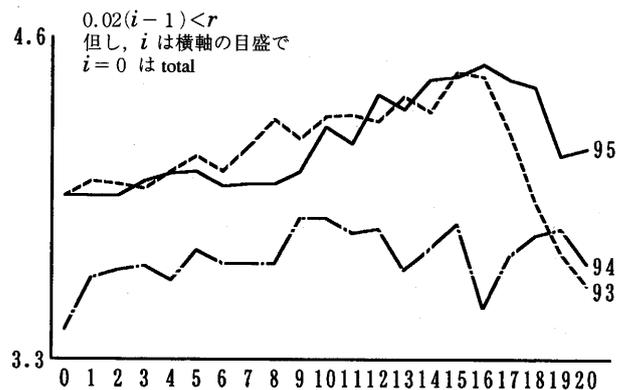


図3-2 rの大きさによって問を制限した場合の z_2 の変化

制限を厳しくするに従って z_2 は次第に大きくなり、 $0.3 < r$ (約70問) よりさらに厳しくすると、これを満たす問の数がかなり少なくなるため、急激に小さくなっている。94年は $0.18 < r$ 辺りにピークがあり、さらに制限を厳しくした場合にも増減を繰り返している。

これに従って、下から10番までに入る不合格者の数は、93年にtotalの場合が5人であるのに対し、 i が14~17の場合はいずれも7人である。94年はtotalの場合が3人であるのに対し、 i が9~15の場合は11と13(各4人)を除いていずれも5人である。また、95年はtotalで不合格者の下からの順位が1, 3, 5で、下から4番までに2人であるのに対し、 i が10~17の場合は11と14(各2人)を除いていずれも3人である。このように、 r がある値より大きい問だけによる点数で順位を付ければ、下から10番まで(95年は4番まで)に入る不合格者の数が2人程度増えるから、このような予測の方法も少なからず役に立つようである。しかし、その年に何人の不合格者が出るかの予測方法も確立しないと、不合格者を高い確率で予測することはできない。

しかしながら、在学中の成績の極めて良くない者が、いわゆる「まぐれ」で国試に合格することは、本来望ましいことではなく、優れた医師を育成するという立場に立てば、下から5人~10人程度を卒延にさせるのも一案であろう。逆に、総合卒試の成績が上位でありながら国

試に不合格となる者がときどき見受けられるが、試験の結果は受験者の精神面や体調などいろいろなコンディションに左右されるから、これについては別の観点から予測しなければならぬ。しかし、それによって卒延にするわけにはいくまい。また、前に述べたように、 t 及び r の小さい間は、問題、正解に不備がなかったとしたら、教育に反省すべき点があるのではないかと考えるなど、教育する者の側に立って、これらのデータを活用すれば、長い目で見て国試の合格率アップに寄与するものと確信する。

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、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三澤恵さんに、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：(1992) 面接の評価・学内成績・医師国家試験の可否の関連。大学入試研究ジャーナル，第2号，58～64
- 2) 平野光昭：(1992) 入学時の平均的学力及び専門教育と医師国家試験の合格率の関連。山梨医科大学紀要，第9巻，84～92
- 3) 平野光昭：(1993) 医師国家試験の合格率を高める要因—受験機会の複数化・入学時の学力レベル・大学教育—。大学入試研究ジャーナル，第3号，23～30
- 4) 平野光昭：(1993) 国立大学の受験機会と入学者の学力レベル及び同レベルと医師国家試験の合格率の関係。大学入学者の特性と選抜方法との関連についての追跡調査研究（平成4年度科学研究費補助金による研究），研究成果報告書，149～156
- 5) 平野光昭：(1993) 卒業試験の成績及び入試成績等と医師国家試験の可否の関係—主成分分析—。山梨医科大学紀要，第10巻，69～78
- 6) 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
- 7) 平野光昭：(1994) 医師国家試験の可否と入学時の属性及び高校調査書の内容の関係—どのような学生を入学させれば国試の合格率が高まるか—。山梨医科大学紀要，第11巻，29～38
- 8) 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，第5号，39～49
- 9) 平野光昭：(1995) 総合卒業試験による医師国家試験の可否予測はどこまで可能か。山梨医科大学紀要，

第12巻，41～49

- 10) 平野光昭：(1996) 追跡調査の理論と実際—追跡調査でこんなにいろいろなことが分かる—。大学入試研究の動向，第11・12合併号，37～56
- 11) 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，第6号，76～83
- 12) 平野光昭，浅香昭雄，北原哲夫：(1996) 推薦選抜における評価の妥当性と信頼性及び同選抜入学者と一般選抜入学者の入学後の成績の比較。大学入試研究ジャーナル，第6号，84～91
- 13) 平野光昭：(1996) 総合卒業試験による医師国家試験の可否予測はどこまで可能か。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年度科学研究費補助金による研究），研究成果報告書，111～120
- 14) 平野光昭，浅香昭雄，北原哲夫：(1997) 推薦選抜における各評価の妥当性と信頼性。大学入試研究ジャーナル，第7号，印刷中
- 15) 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

Abstract**How Far Is It Possible for Any Student to Predict the Pass-rate of
National Examination for Medical License (NE)
with the Aid of the Comprehensive
Graduation Test (CGT) ?
(Part 2)****Teruaki HIRANO**

The elevation of the quality of successful applicants and the improvement of the contents of education result in the increase of the number of medical students who are to go straight on to doctors in due course of six years and the birth of promising doctors. 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, applying the theory of the item response and using several kinds of standard selected from CGT particular questions which seem to serve as the decisive factors to distinguish between the able student and the unable one. Then we have been trying to find with the aid of the particular scores in these questions how far it is possible for any student to predict the pass-rate of NE. In this essay, the author tried to improve upon the method of analysis using the data on the students who graduated between '94 and '96, and it became possible to enhance the power of prediction more effectively than doing it by the method in the previous essay.

Department of Mathematics