

大学入試センター試験得点の標準化の試み

—大学入試センター 前川 眞一・石塚 智一・菊地 賢一・内田 照久・中畝 菜穂子—

1. はじめに

大学入試センターが実施してきたテスト(共通第1次学力試験, 大学入試センター試験)は, 毎年新しい問題が作成されており, 受験生も重複受験者を除けば毎年新しくなる。したがって, 各テストの平均点は, 年度ごとの問題の難易度と受験生の学力の二つの要因からの影響を受け, その差異が直接問題の難易度や受験生の学力の差異を示すものではない。平均点の差は, 受験生の学力は不変であることを仮定すれば, 問題の難易度の差に, また, 問題の難易度が不変であることを仮定すれば, 学力の差であると解釈できるが, これらの仮定が正しいという保証はない。また, 重複受験者のデータのみから推論を行うことも可能であるが, 重複受験者の大部分は学力がのびている可能性があるため, この方法もあまり有効ではない。

本研究では, 実験的に過去の問題を被験者に解答させ, そのデータを基に過去の試験の難易度の推定を行なう。理想的には, 被験者が過去20年間の試験問題冊子に含まれる問題の全てに解答することが望ましいが, 現実問題としてそれは不可能であるため, 過去の試験問題冊子からいくつかの大問をサンプルして新たな冊子を複数作り, それらを被験者に解答させる。その際, 新たな冊子が全く別々の問題を含むのではなく, いくつかの大問が複数の冊子に含まれるようにすることにより, 全大問のリンクがとれるようにする。また, 同時に, 古典的テスト理論のモデルを用いた分析を行うことにより, 各年度ごとの受験者層の学力の推移を推定する。

2. 実験のデザイン

ある科目の n 年分の冊子に含まれる大問を A と B の 2 群に分け, A 群には各年度から同数 (d 個) の大問が含まれるものとする。A 群から新たな冊子を n 冊作り, 各冊子には, m 年分の大問がすべて含まれるようにする。また, 大問間のリンクを取るために, 2 冊子間で重複する年度を r とする。

以下の表の左上部に $n = 5, d = 2, m = 3, r = 2$ の場合の例を示すが, 各年度の大問総数を 4 とし, そのうち A 群として第 1 問と第 2 問を, B 群として第 3 問と第 4 問を用いている。なお, 各冊子 $Ge_1 - Ge_5$ は, 表中 xx で示した部分の大問のみを含んでいる。

year	Group A					Group B				
	Y1	Y2	Y3	Y4	Y5	Y1	Y2	Y3	Y4	Y5
daimon	12	12	12	12	12	34	34	34	34	34
daimon ID	DD	DD	DD	DD	D	DD	DD	DD	DD	DD
	12	34	56	78	D1	12	34	56	78	90
group										
Ge1	xx	xx	xx
Ge2	..	xx	xx	xx
Ge3	xx	xx	xx
Ge4	xx	xx	xx
Ge5	xx	xx	xx
G1
G2
G3	xx	xx
G4	xx	xx	..
G5	xx	xx

図1 実験デザインの概念図

例えば, 第 1 の冊子は第 1, 2, 3 年度の $m = 3$ 年分の大問(D1-D6)を, また, 第 2 の冊子は第 2, 3, 4 年度の 3 年分の大問(D3-D8)を含むことになり, 第 1 の冊子と第 2 の冊子は第 2, 3 年度の $r = 2$ 年度の大問(D3-D6)を共有している。同様に, 第 4 の冊子は第 1, 4, 5 年度の 3 年度の大問を, また, 第 5 の冊子は第 1, 2, 5 年度の 3 年度の大問を含むことになり, 第 4 の冊子と第 5 の冊子は第 1, 5 年度の 2 年度の大問(D1 D2 D9 D10)を共有している。なお, 各大問は, それぞれ 3 つの冊子に含まれ

ている。

これらの冊子をそれぞれ等質であると考えられる被験者群に解答させる。これらの被験者群を実験群と呼び $G_i, i = 1, 2, \dots, n$ で表す。また、これらの実験データに加えて、実際の試験の実施結果から実受験者群 $G_i, i = 1, 2, \dots, n$ のデータを得る。

上記の例では、実験に含まれなかった大問が各年度ごとに2問ずつあるが、上記の表の下部に実受験者群の冊子の内容を、B群の大問をも含む形で示した。

各被験者群の人数を n_k とし、上記の各セルの xx の部分に入る観測値を $y_{ij}^{(k)}$ とする。なお、以下の表記では、添字 i, j, k は、それぞれ集団内の個人、各大問、集団を表すものとする。

3. 古典的テスト理論のモデル

3-1 同族テストモデル

古典的テスト理論では、観測される複数のテスト得点とその真値の間に様々な仮定をもうけているが、同族テスト (Congeneric Tests) と呼ばれるものはその中でもかなり一般的なものである。本節の議論では、図1に示した各大問をあたかもそれ自身一つのテストであるとして取り扱う。なお、古典的テスト理論に関しては、例えば、池田(1973)を参照されたい。

いま、 N 人が p 個のテストを受験しているとすれば、同族テストの仮定は

$$y_{ij} = \mu_j + t_i \lambda_j + \varepsilon_{ij} \quad (1)$$

ただし、 $y_{ij}, i = 1, 2, \dots, N, j = 1, 2, \dots, p$ は個人 i のテスト(ここでは大問) j における得点であり、 μ_j はテスト j の難易度(正確には値が大きいほど易しいという非困難度)を表すパラメタ、また、 t_i は全てのテストが測定しようとしている共通の真値、そして ε_{ij} は誤差項である。また、 λ_j は正の係数で、テストの識別力を表すものである。また、誤差項に関し

ては、その平均は0とし、分散を σ_j^2 と置く。

なお、パラメタの識別性のための制約条件

$$E(t_i) = 0 \quad \text{and} \quad \text{var}(t_i) = 1 \quad (2)$$

を置く。これは、個人の真値の母集団における分布の平均と分散を定めているもので、

$$E(y_{ij} | t_i) = \mu_j + t_i \lambda_j \quad (3)$$

であり、

$$E(y_{ij}) = \mu_j \quad (4)$$

となる。また、分散に関しては、真値と誤差は独立であるため、

$$\text{var}(y_{ij} | t_i) = \sigma_j^2 \quad \text{and} \quad \text{var}(y_{ij}) = \lambda_j^2 + \sigma_j^2 \quad (5)$$

である。

なお、(1)式と(2)式に示される同族テストのパラメタに付加的な制約条件を課したものには様々な名前が付いているが、ここでは本質的にタウ等価なテストに関して説明する。これは同族テストにおいて

$$\lambda_j = \lambda, j = 1, 2, \dots, p \quad (6)$$

とおいたものであるが、通常は

$$y_{ij} = \mu_j + t_i + \varepsilon_{ij} \quad (7)$$

ただし、

$$E(t_i) = 0 \quad (8)$$

という形で書かれることが多い。この場合、真値の分散を σ_j^2 と記す。なお、これは

$$\sum_{j=1}^p \mu_j = 0 \quad (9)$$

という制約と本質的に同値である。

(7)式が示すように、本質的にタウ等価なテストのモデルは、その誤差分散に関する仮定を除けば、反復測定を含む1元配置モデルに等しい。また、次節に述べるように、各集団の母平均が異なると考えられる場合にはひとつの被験者間要因とひとつの被験者内要因を持つ(反復測定を含む)分散分析モデルとなる。

3-2 モデルパラメタの推定

図 1 のデータの各セルがすべて観測されていれば、そのデータを因子分析する事により、同族テストのパラメタ $\mu_j, \lambda_j, \sigma_j^2, j=1, 2, \dots, p$ を求めることができる。しかし、実際には、テスト j が冊子 k に含まれない場合は $y_{ij}^{(k)}$ は欠測となっている。また、それぞれの冊子が異なる G 個の集団で解答されており、それらの集団の等質性が必ずしも保証されていない。(図 1 の例では、実験群は全て等質であるかもしれないが、実受験者群は等質である保証はない。)

この様なデータから同族テストのパラメタを推定するための方法はいろいろあるが、ここでは以下のような仮定の下に市販の構造方程式モデル(例えば豊田, 1998 を参照)のプログラムを利用することを考える。

1. λ は集団に依存しない。
2. μ は集団に依存しない。
3. 真値の平均と分散は集団ごとに異なる。
4. 誤差分散は集団に依存しない。

従って、この3つの仮定の下では、

$$y_{ij}^{(k)} = \mu_j + \lambda_j t_i^{(k)} + \varepsilon_{ij}^{(k)}, \quad i=1, 2, \dots, N, \\ j=1, 2, \dots, p, k=1, 2, \dots, G \quad (10)$$

ただし、

$$E(t_i^{(k)}) = \xi_{(k)} \quad \text{and} \\ \text{var}(t_i^{(k)}) = \phi_{(k)}, \quad k=1, 2, \dots, G \quad (11)$$

$$E(\varepsilon_{ij}^{(k)}) = 0 \quad \text{and} \\ \text{var}(\varepsilon_{ij}^{(k)}) = \sigma_j^2, \quad k=1, 2, \dots, G \quad (12)$$

という形でパラメタを求めればよい。ただ、原点と単位の不定性を取り除くために、 $\xi_{(1)}=0, \phi_{(1)}=1$ 等の制約を置く。

タウ等価モデルを仮定した場合には

$$\lambda_j = \lambda \quad (13)$$

という制約をかければよい。

いずれにしても、各集団ごとに観測される

変数のセットが異なり、因子の平均と分散を集団ごとに推定することが可能な構造方程式モデルのプログラムが必要となる。(複数母集団でシステマティックな変数単位の欠測データを含むデザインである。)

3-3 数値例

図 1 のデザインを用いて、実験群に 40 名、実受験者群に 1000 名の受験者がいる状況をシミュレートした。ただし、B 群の大問も各年度 2 個とし、5 年間全体で 20 個の大問(D1-D20)が存在するものとしている。すなわち、大問 1 から 10 までが実験に用いられ、大問 11 から 20 までが実受験者群でのみ用いられている。1 因子の同族テストの仮定の下にデータを発生させ、それから、Amos 4 (Arbuckle, 1999)を用いて元のパラメタがどのように復元されるかを調べた。データの生成に用いたパラメタの値と推定値を表 1 に示す。実際の実験では、実験群の学力はバランスがとれているはずであるため、それらの平均と分散を 0 と 1 に固定した。また、誤差分散は共通とした。実受験者群には、実際には 500,000 程度の受験者がいるが、ここでは 1000 人としている。

表 1 True and Estimated Parameters

	μ		λ		σ^2	
	True	Estimated	True	Estimated	True	Estimated
D1	60.000	60.477	9.000	8.811	18.000	15.983
D2	80.000	80.856	7.000	7.010	18.000	16.791
D3	70.000	70.074	9.000	8.311	18.000	15.212
D4	75.000	75.483	7.000	6.379	18.000	16.858
D5	60.000	60.844	9.000	8.304	18.000	15.032
D6	60.000	60.576	7.000	6.318	18.000	15.509
D7	50.000	51.130	9.000	8.712	18.000	14.415
D8	45.000	45.637	7.000	6.694	18.000	15.485
D9	60.000	60.720	9.000	9.082	18.000	15.854
D10	60.000	60.425	7.000	7.047	18.000	14.873
D11	60.000	60.761	7.000	7.052	18.000	16.026
D12	60.000	60.767	9.000	8.948	18.000	15.522
D13	70.000	70.353	7.000	6.514	18.000	16.199
D14	75.000	75.422	9.000	9.463	18.000	17.323
D15	60.000	60.486	7.000	6.430	18.000	15.091
D16	60.000	60.814	9.000	8.347	18.000	16.211
D17	50.000	50.205	7.000	6.394	18.000	16.128
D18	45.000	45.741	9.000	8.385	18.000	16.880
D19	60.000	60.526	7.000	7.269	18.000	14.031
D20	60.000	60.767	9.000	8.657	18.000	16.377

		ξ		ϕ	
		True	Estimated	True	Estimated
Group1	Ge1	0.000	0.000	1.000	1.000
Group2	Ge2	0.000	0.000	1.000	1.000
Group3	Ge3	0.000	0.000	1.000	1.000
Group4	Ge4	0.000	0.000	1.000	1.000
Group5	Ge5	0.000	0.000	1.000	1.000
Group6	G1	0.000	-0.108	1.000	1.017
Group7	G2	1.000	1.019	1.000	1.163
Group8	G3	0.000	-0.052	0.640	0.730
Group9	G4	-1.000	-1.141	0.640	0.730
Group10	G5	0.000	-0.082	0.490	0.531

推定に際しては、実験群の学力はその平均と分散を0と1に固定し、実受験者群のそれは自由パラメタとした。また、誤差分散は全てのものを自由に推定した。

誤差分散をかなり大きめにしているにもかかわらず、パラメタはかなりうまく推定され

ていることが分かる。特に、観測変数に係わるパラメタ μ, λ, σ^2 の推定の方が、潜在変数のパラメタ ξ, ϕ に比してその精度が良いという印象を受ける。なお、 μ の推定値は全て真の値を上回っているが、この理由は定かではない。

表2 問題冊子(英語)

冊子1	版1-1	S56-2,4,5	S61-5,6,7	H4-3A,3B,5
	版1-2	S61-5,6,7	H4-3A,3B,5	S56-2,4,5
	版1-3	H4-3A,3B,5	S56-2,4,5	S61-5,6,7
冊子2	版2-1	S61-5,6,7	H4-3A,3B,5	H8-3A,3B,5
	版2-2	H4-3A,3B,5	H8-3A,3B,5	S61-5,6,7
	版2-3	H8-3A,3B,5	S61-5,6,7	H4-3A,3B,5
冊子3	版3-1	H4-3A,3B,5	H8-3A,3B,5	S59-5,7,9
	版3-2	H8-3A,3B,5	S59-5,7,9	H4-3A,3B,5
	版3-3	S59-5,7,9	H4-3A,3B,5	H8-3A,3B,5
冊子4	版4-1	H8-3A,3B,5	S59-5,7,9	H3-3,5
	版4-2	S59-5,7,9	H3-3,5	H8-3A,3B,5
	版4-3	H3-3,5	H8-3A,3B,5	S59-5,7,9
冊子5	版5-1	S59-5,7,9	H3-3,5	H7-3A,3B,5
	版5-2	H3-3,5	H7-3A,3B,5	S59-5,7,9
	版5-3	H7-3A,3B,5	S59-5,7,9	H3-3,5
冊子6	版6-1	H3-3,5	H7-3A,3B,5	S57-5,8,9
	版6-2	H7-3A,3B,5	S57-5,8,9	H3-3,5
	版6-3	S57-5,8,9	H3-3,5	H7-3A,3B,5
冊子7	版7-1	H7-3A,3B,5	S57-5,8,9	H1-4,5,8
	版7-2	S57-5,8,9	H1-4,5,8	H7-3A,3B,5
	版7-3	H1-4,5,8	H7-3A,3B,5	S57-5,8,9
冊子8	版8-1	S57-5,8,9	H1-4,5,8	H5-3A,3B,5
	版8-2	H1-4,5,8	H5-3A,3B,5	S57-5,8,9
	版8-3	H5-3A,3B,5	S57-5,8,9	H1-4,5,8
冊子9	版9-1	H1-4,5,8	H5-3A,3B,5	S56-2,4,5
	版9-2	H5-3A,3B,5	S56-2,4,5	H1-4,5,8
	版9-3	S56-2,4,5	H1-4,5,8	H5-3A,3B,5
冊子10	版10-1	H5-3A,3B,5	S56-2,4,5	S61-5,6,7
	版10-2	S56-2,4,5	S61-5,6,7	H5-3A,3B,5
	版10-3	S61-5,6,7	H5-3A,3B,5	S56-2,4,5

表3 問題冊子(数学)

冊子1	版1-1	S56-1,3	S61-1,4	H4-I-1,II-1
	版1-2	S61-1,4	H4-I-1,II-1	S56-1,3
	版1-3	H4-I-1,II-1	S56-1,3	S61-1,4
冊子2	版2-1	S61-1,4	H4-I-1,II-1	H8-II-1
	版2-2	H4-I-1,II-1	H8-II-1	S61-1,4
	版2-3	H8-II-1	S61-1,4	H4-I-1,II-1
冊子3	版3-1	H4-I-1,II-1	H8-II-1	S59-2,3
	版3-2	H8-II-1	S59-2,3	H4-I-1,II-1
	版3-3	S59-2,3	H4-I-1,II-1	H8-II-1
冊子4	版4-1	H8-II-1	S59-2,3	H3-I-3,II-2
	版4-2	S59-2,3	H3-I-3,II-2	H8-II-1
	版4-3	H3-I-3,II-2	H8-II-1	S59-2,3
冊子5	版5-1	S59-2,3	H3-I-3,II-2	H7-II-2
	版5-2	H3-I-3,II-2	H7-II-2	S59-2,3
	版5-3	H7-II-2	S59-2,3	H3-I-3,II-2
冊子6	版6-1	H3-I-3,II-2	H7-II-2	S57-2,4
	版6-2	H7-II-2	S57-2,4	H3-I-3,II-2
	版6-3	S57-2,4	H3-I-3,II-2	H7-II-2
冊子7	版7-1	H7-II-2	S57-2,4	H1-3,4
	版7-2	S57-2,4	H1-3,4	H7-II-2
	版7-3	H1-3,4	H7-II-2	S57-2,4
冊子8	版8-1	S57-2,4	H1-3,4	H5-I-3,II-1
	版8-2	H1-3,4	H5-I-3,II-1	S57-2,4
	版8-3	H5-I-3,II-1	S57-2,4	H1-3,4
冊子9	版9-1	H1-3,4	H5-I-3,II-1	S56-1,3
	版9-2	H5-I-3,II-1	S56-1,3	H1-3,4
	版9-3	S56-1,3	H1-3,4	H5-I-3,II-1
冊子10	版10-1	H5-I-3,II-1	S56-1,3	S61-1,4
	版10-2	S56-1,3	S61-1,4	H5-I-3,II-1
	版10-3	S61-1,4	H5-I-3,II-1	S56-1,3

実験冊子1	S56	S61	H4						
実験冊子2		S61	H4	H8					
実験冊子3			H4	H8	S59				
実験冊子4				H8	S59	H3			
実験冊子5					S59	H3	H7		
実験冊子6						H3	H7	S57	
実験冊子7							H7	S57	H1
実験冊子8								S57	H1
実験冊子9	S56								H1
実験冊子10	S56	S61							H5
本番冊子1	S56								
本番冊子2		S61							
本番冊子3			H4						
本番冊子4				H8					
本番冊子5					S59				
本番冊子6						H3			
本番冊子7							H7		
本番冊子8								S57	
本番冊子9									H1
本番冊子10									H5

図2 実験デザイン

4. 実験データの解析

実際の実験では、試験科目として英語と数学の二つを用いた。

4-1 実験のデザイン

昭和 56, 57, 59, 61 年度と平成元年度の共通第 1 次学力試験、ならびに平成 3, 4, 5, 7, 8 年度の大学入試センター試験の 10 年度にわたる数学と英語の問題から 25 題の大問を選出し、それらを実験で用いる A 群とした。この際、各年度から少なくとも 2 つの大問が選ばれるようにした。これらの大問を組み合わせ、各問題冊子が必ず 3 つの年度の問題を全て含むように、また各年度の問題が必ず 3 つの冊子に含まれるように、問題冊子を編集した。その後、問題冊子の中の大問の出現位置の影響を除去するため、各冊子から 3 つの版を作成し、それぞれで各年度に属する大問の出現場所を変化させた。したがって、各年度に属する大問は、問題冊子の前半、中盤、後半に同頻度で出現することとなった。その内訳を、表 2, 表 3 と図 2 に示す。

被験者としては、3 つの国立大学から集めた約 400 人の大学 1 年生(平成 11 年度)を用いたが、被験者は各問題冊子で均等となるようにランダムに割り振ったため、各問題冊子は約 40 名(各版はその 1/3)の被験者に割り振られたことになる。その結果、各大問は約 120 人の被験者に受験された。

4-2 実験結果 I

各冊子の合計得点を大学別に標準化して、2 標準偏差より低い得点の受験者やそうでなくても特に得点の低かった受験者について、問題用紙と答案用紙を確認した。その結果、本来の実力通りに試験を受けていないものと判断された受験者に関しては、データ解析の対象外とした。この際、当日の受験態度や数学と英語双方の得点をつき合わせた結果も考

慮した。その結果、英語に関しては 13 名を、数学に関しては 16 名を対象外とした。

この、スクリーニングを経たデータの平均得点率(配点を考慮して採点した大問得点をその満点で除したも)等を表 4 と表 5 に示す。また、同表の右側に、実受験者群(実際の実施結果)で得られた平均得点率等を示した。

表 4 英語平均得点率等

	実 験 群			実 受 験 者 群		
	N	Mean	Std Dev	N	Mean	Std Dev
1 56-2	117	0.75	0.19	336623	0.59	0.26
2 56-4	117	0.66	0.24	336623	0.40	0.27
3 56-5	117	0.72	0.31	336623	0.53	0.34
4 57-5	120	0.75	0.15	330672	0.53	0.23
5 57-8	120	0.80	0.24	330672	0.62	0.33
6 57-9	120	0.77	0.27	330672	0.54	0.31
7 59-5	118	0.70	0.24	338323	0.52	0.29
8 59-7	118	0.73	0.28	338323	0.56	0.32
9 59-9	118	0.71	0.20	338323	0.58	0.24
10 61-5	118	0.79	0.23	349533	0.52	0.31
11 61-6	118	0.59	0.32	349533	0.44	0.37
12 61-7	118	0.60	0.29	349533	0.42	0.30
13 1-4	116	0.73	0.22	379990	0.54	0.27
14 1-5	116	0.68	0.27	379990	0.46	0.30
15 1-8	116	0.69	0.29	379990	0.50	0.32
16 3-3	120	0.58	0.25	428564	0.44	0.28
17 3-5	120	0.74	0.24	428564	0.57	0.29
18 4-3	119	0.74	0.23	443730	0.50	0.28
19 4-5	119	0.81	0.27	443730	0.59	0.34
20 5-3	115	0.71	0.21	479095	0.42	0.27
21 5-5	115	0.70	0.26	479095	0.49	0.27
22 7-3	119	0.74	0.21	517351	0.47	0.28
23 7-5	119	0.64	0.28	517351	0.45	0.31
24 8-3	119	0.76	0.20	530734	0.56	0.29
25 8-5	119	0.70	0.30	530734	0.52	0.33

表 5 数学平均得点率等

	実 験 群			実 受 験 者 群		
	N	Mean	Std Dev	N	Mean	Std Dev
1 56-1	118	0.73	0.25	340578	0.60	0.30
2 56-3	118	0.76	0.32	340578	0.58	0.39
3 57-2	116	0.76	0.30	334102	0.52	0.33
4 57-4	116	0.70	0.28	334102	0.52	0.29
5 59-2	117	0.77	0.30	341375	0.53	0.35
6 59-3	117	0.83	0.20	341375	0.56	0.32
7 61-1	116	0.60	0.26	345848	0.59	0.30
8 61-4	116	0.56	0.40	219410	0.33	0.38
9 1-3	117	0.67	0.26	383324	0.47	0.30
10 1-4	117	0.75	0.23	332750	0.53	0.30
11 3-I-3	119	0.68	0.22	330228	0.61	0.31
12 3-II-2	119	0.68	0.26	268122	0.64	0.28
13 4-I-1	115	0.76	0.29	336835	0.54	0.32
14 4-II-1	115	0.62	0.33	293752	0.49	0.27
15 5-I-3	117	0.82	0.21	351656	0.64	0.28
16 5-II-1	117	0.74	0.25	303368	0.65	0.28
17 7-II-2	117	0.77	0.21	300012	0.69	0.26
18 8-II-1	115	0.64	0.27	327114	0.57	0.27

各問題冊子を受験した被験者群間でその学力に差がないことを調べるために、問題冊子とその中に入れ子になった版を被験者間要因、大問を被験者内要因とした 3 元配置の反復測定のある分散分析モデルで分析した結果、問題冊子の効果は 5% で有意ではないという結果を得た。

また、出現位置の効果を見るために、各年度に属する大問の平均得点率の平均を、出現位置を要因とする 1 元配置の分散分析で分析した結果、出現位置の効果は 5% で有意ではないという結果を得た。

表6 同族テストモデルのパラメタの推定値

英語	入	σ^2	μ	*	数学				
					μ	入	σ^2	*	
H1811-	0.670	0.072	0.008	*	D2811-	0.735	0.170	0.043	*
H1812-	0.761	0.100	0.040	*	D2812-	0.762	0.159	0.039	*
H1813-	0.829	0.089	0.024	*	D2813-	0.751	0.210	0.077	*
H1814-	0.636	0.133	0.035	*	D2814-	0.748	0.173	0.058	*
H1815-	0.727	0.111	0.089	*	D2815-	0.785	0.178	0.074	*
H1816-	0.590	0.128	0.046	*	D2821-	0.642	0.143	0.042	*
H1817-	0.593	0.119	0.074	*	D2822-	0.718	0.227	0.060	*
H1818-	0.682	0.107	0.043	*	D2823-	0.914	0.111	0.030	*
H1821-	0.680	0.089	0.015	*	D2825-	0.751	0.161	0.044	*
H1822-	0.938	0.092	0.039	*	D2826-	0.652	0.137	0.052	*
H1823-	0.773	0.092	0.046	*	D2841-	0.888	0.135	0.052	*
H1824-	0.716	0.120	0.034	*	D2842-	0.801	0.132	0.051	*
H1825-	0.742	0.116	0.022	*	D2843-	0.819	0.150	0.045	*
H1826-	0.924	0.116	0.080	*	D2844-	0.579	0.150	0.111	*
H1827-	0.740	0.134	0.106	*	D2845-	0.826	0.202	0.046	*
H1828-	0.841	0.130	0.062	*	DE861-	0.823	0.164	0.097	*
H1829-	0.785	0.137	0.052	*	DE862-	0.853	0.168	0.047	*
H18210	0.739	0.120	0.043	*	DE863-	0.712	0.168	0.065	*
H1841-	0.589	0.079	0.035	*	DE864-	0.522	0.191	0.080	*
H1842-	0.692	0.092	0.031	*	DE865-	0.719	0.217	0.076	*
H1843-	0.773	0.131	0.032	*	DE866-	0.800	0.124	0.080	*
H1844-	0.605	0.102	0.030	*	DE891-	0.895	0.147	0.047	*
H1845-	0.684	0.119	0.055	*	DE892-	0.855	0.153	0.047	*
H1846-	0.815	0.064	0.064	*	DE893-	0.692	0.171	0.046	*
H1847-	0.728	0.124	0.069	*	DE894-	0.763	0.176	0.039	*
H1848-	0.775	0.111	0.052	*	DE895-	0.823	0.138	0.050	*
H1849-	0.717	0.107	0.033	*	DE896-	0.309	0.058	0.028	*
H1861-	0.877	0.074	0.023	*	D2811-	0.652	0.058	0.032	*
H1862-	0.842	0.109	0.002	*	D2812-	0.515	0.151	0.042	*
H1863-	0.872	0.129	0.015	*	D2813-	0.712	0.156	0.062	*
H1864-	0.984	0.096	0.015	*	D2814-	0.662	0.156	0.034	*
H1865-	0.761	0.135	0.039	*	D2815-	0.788	0.182	0.039	*
H1866-	0.628	0.117	0.106	*	D2816-	0.689	0.123	0.045	*
H1867-	0.610	0.114	0.085	*	D2821-	0.741	0.187	0.043	*
H1891-	0.785	0.085	0.016	*	D2822-	0.833	0.100	0.035	*
H1892-	0.682	0.094	0.021	*	D2823-	0.657	0.157	0.047	*
H1893-	0.866	0.088	0.016	*	E2921-	0.599	0.133	0.030	*
H1894-	0.731	0.134	0.035	*	E2922-	0.622	0.112	0.051	*
H1895-	0.674	0.147	0.044	*	D2931-	0.834	0.094	0.026	*
H1896-	0.664	0.127	0.048	*	D2932-	0.599	0.163	0.038	*
H1897-	0.621	0.101	0.053	*	D2933-	0.779	0.159	0.039	*
H1898-	0.570	0.115	0.074	*	D2934-	0.771	0.159	0.036	*
H1899-	0.529	0.096	0.032	*	E2931-	0.855	0.178	0.043	*
H1911-	0.693	0.094	0.032	*	E2932-	0.675	0.192	0.070	*
H1912-	0.773	0.104	0.008	*	D2951-	0.888	0.118	0.042	*
H1913-	0.595	0.136	0.038	*	D2952-	0.619	0.175	0.048	*
H1914-	0.967	0.054	0.035	*	D2953-	0.445	0.102	0.027	*
H1915-	0.724	0.137	0.043	*	E2954-	0.751	0.156	0.051	*
H1916-	0.957	0.122	0.032	*	E2955-	0.719	0.149	0.032	*
H1921-	0.762	0.082	0.027	*	D2961-	0.719	0.159	0.052	*
H1922-	0.806	0.095	0.032	*	D2962-	0.685	0.108	0.050	*
H1923-	0.609	0.140	0.017	*	D2963-	0.765	0.106	0.050	*
H1924-	0.714	0.077	0.031	*	D2964-	0.564	0.086	0.022	*
H1925-	0.847	0.077	0.059	*	D2965-	0.492	0.152	0.068	*
H1926-	0.673	0.099	0.029	*	E2961-	0.650	0.152	0.041	*
H1931-	0.837	0.092	0.028	*	E2962-	0.557	0.139	0.025	*
H1932-	0.752	0.093	0.011	*	E2963-	0.538	0.083	0.067	*
H1933-	0.706	0.083	0.011	*					
H1934-	0.926	0.131	0.034	*					
H1935-	0.717	0.109	0.063	*					
H1936-	0.745	0.108	0.049	*					
H1937-	0.679	0.108	0.032	*					
H1951-	0.579	0.076	0.030	*					
H1952-	0.814	0.096	0.012	*					
H1953-	0.748	0.136	0.036	*					
H1954-	0.771	0.108	0.071	*					
H1955-	0.558	0.103	0.074	*					
H1956-	0.713	0.106	0.026	*					
H1959-	0.773	0.105	0.026	*					
H1961-	0.787	0.103	0.033	*					
H1962-	0.793	0.098	0.010	*					
H1963-	0.761	0.149	0.029	*					
H1964-	0.981	0.149	0.044	*					
H1965-	0.981	0.072	0.072	*					
H1966-	0.694	0.128	0.073	*					
H1966-	0.774	0.135	0.024	*					

表中 * 印は A 群に含まれる大問を示す。

表7 得点率・パラメタ等の年度別平均

英語	入	MEAN_E1	NEAN1	NEAN2	英語	入	MEAN_E1	NEAN1	NEAN2
56 81:	0.739	0.709	0.553	0.510	56 81:	0.749	0.754	0.610	0.594
57 82:	0.783	0.775	0.595	0.548	57 82:	0.756	0.736	0.623	0.573
59 84:	0.707	0.713	0.566	0.548	59 84:	0.771	0.805	0.527	0.549
61 86:	0.810	0.691	0.632	0.488	61 86:	0.725	0.697	0.585	0.517
1 89:	0.711	0.692	0.583	0.504	1 89:	0.806	0.719	0.505	0.501
3 91:	0.750	0.657	0.634	0.508	3 91:	0.712	0.785	0.592	0.633
4 92:	0.808	0.774	0.644	0.548	4 92:	0.700	0.692	0.546	0.519
5 93:	0.787	0.704	0.557	0.456	5 93:	0.814	0.786	0.695	0.648
7 95:	0.734	0.689	0.528	0.461	7 95:	0.698	0.776	0.621	0.699
8 96:	0.797	0.731	0.652	0.544	8 96:	0.805	0.641	0.536	0.572

INT1 μ の平均
 MEAN_E1 A 群の大問得点率の平均 (受験者群)
 NEAN1 平均点 / 満点 (受験者群)
 NEAN2 A 群の大問得点率の平均 (受験者群)

4-3 実験結果 II

本節では、実験データの古典的テストモデルによる分析結果を示す。前半に述べたように、古典的テストモデルを用いた分析は、実験に用いられた大問(A群)のみを用いても、全ての大問を用いても行うことが出来るが、ここでは全ての大問を用いた分析結果を示す。なお、分析に際して、実受験者群のデータとしては全受験者を用いて計算された分散共分散行列を用いたが、そのサンプルサイズを

500 とした。また、計算は前節のシミュレーションと同様 Amos4 を用いたが、 ϕ パラメータは、実受験者群で共通とした。

表 6 に同族テストモデルのパラメータの推定値を示す。また、表 7 に、推定された大問の難易度の年度ごとの平均、A群に属する大問得点率の実験群並びに実受験者群における年度ごとの平均、並びに、全受験者を用いて計算された基準化平均点、すなわち、所謂平均点(A-B両群を併せたもの)を満点で除したものを示す。これらの年度ごとの平均をプ

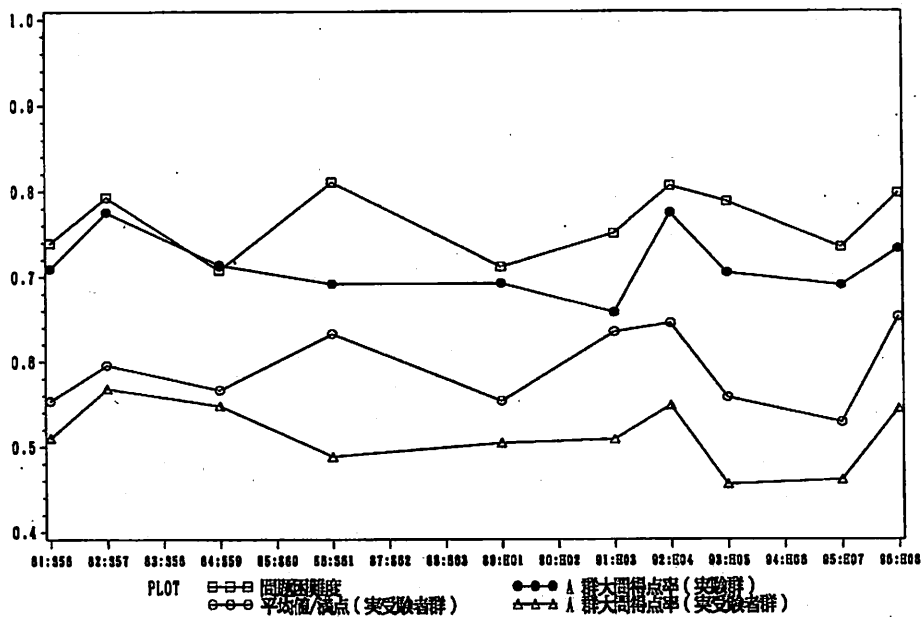


図3 年度ごとの問題の難易度 (英語)

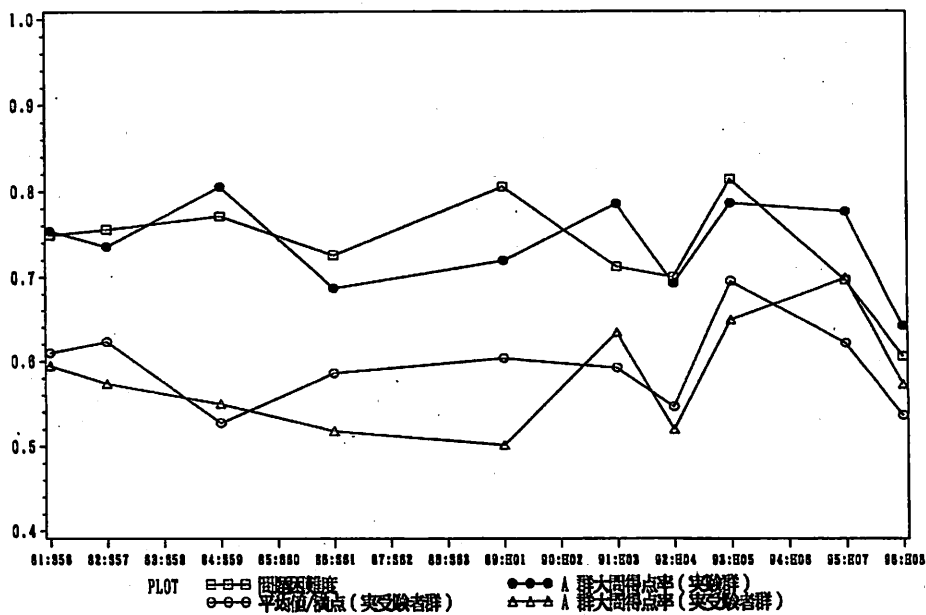


図4 年度ごとの問題の難易度 (数学)

ロットしたものが図3と図4であるが、プロットにおいて顕著な時系列的な傾向性は見あたらない。

A群における大問得点率の二つのプロットは、同じ問題を異なる受験者群に課した際の得点率の平均値であるから、ほぼ等間隔(平行)になっている。また、推定された大問の難易度と、基準化平均点もほぼ等間隔になっている。この傾向は、特に英語において顕著である。なお、これらの2本のプロットの差(実験群における平均-実受験者群における平

均、もしくは、推定された大問の難易度の平均-基準化平均点)が受験者の学力を反映していると考えられるが、そのことを図5と図6に示した。

図5と図6は、同族テストモデルによる分析で得られた実受験者群の学力((11)式の k (k)の推定値と図3と図4における二つのプロットの差を同時に示したものである。なお、縦軸の尺度は、実験群の学力の平均値を0、分散を1にし固定したものになっているため、推定された学力は全て負の値となっている。

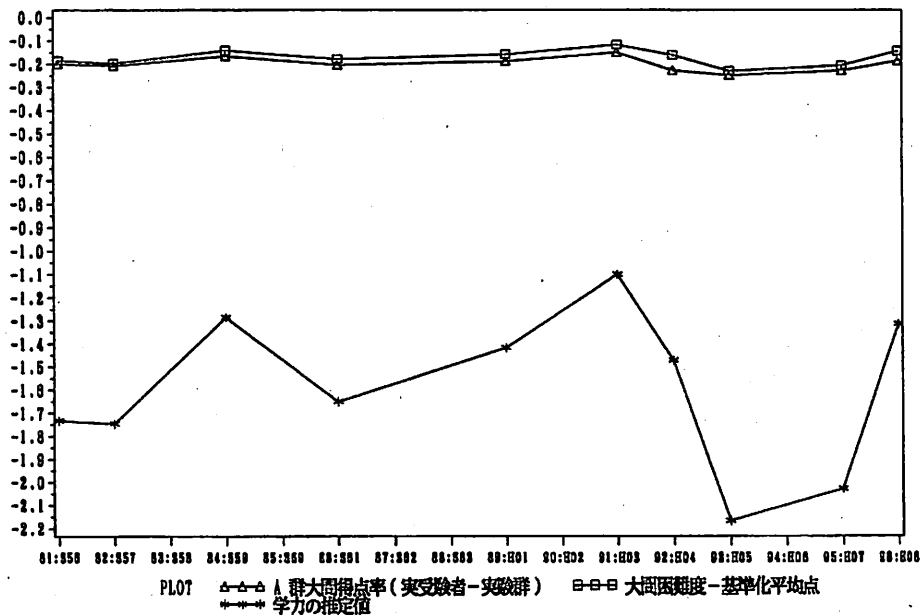
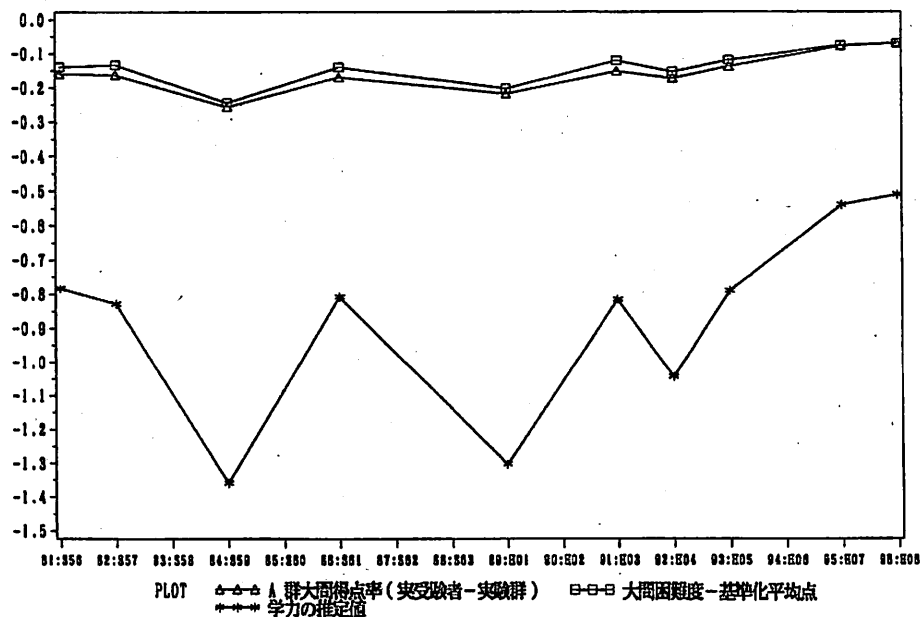


図5 実受験者群の学力の推定値 (英語)



(実験に使用した被験者集団は、全受験者集団と比較して学力がかなり高い方に位置している。)また、この推定値の標準誤差はほぼ0.15-0.20の間に入っている。

これから、 $\xi_{(a)}$ はA群に属する大問得点率の実験群と実受験者群における差異、同時に、同族テストモデルによって推定された大問の難易度と、基準化された平均点の差異を反映したものとなっていることが分かる。

また、学力の推移の時系列的傾向は、数学においてわずかな上昇傾向が見られるが、ほとんどランダムであるように見える。

5. 結果と考察

本研究では、過去の共通第1次学力試験と大学入試センター試験の中から選ばれた問題を用いて、各年度に含まれる問題の難易度と各年度の受験生の学力を実験的に推定することを試みた。その結果、問題の難易度に関しても、受験生の学力に関しても、その時系列的变化に明らかな傾向を見いだすことは出来なかった。

分析の方法としては、古典的テスト理論の同族テストモデルを用いたが、今回の実験におけるその有効性は、シミュレーションならびに、推定された学力のプロットから示されたと考えられる。特に、学力の推定値と、実験群と実受験者群における平均得点率の差異がほぼ同様の傾向を示すことは、このアプローチの健全さを物語っているといえる。また、学力の推定値が、同族テストモデルによって推定された大問の難易度と、基準化平均点の差異を反映したものとなっていることは、実験に含まれなかったB群の大問の難易度もこのアプローチにより推定可能であることを示すものである。なぜなら、基準化平均点にはB群の大問も寄与しているからである。ただ、今回の実験において、このような比較が可能となったのは、バランスのとれた実験デザインを用いたからであるが、同族テスト

モデルによる分析はバランスの取れていない場合にも適用可能である。今後も、同様の実験を行い、データ数を増やすとともに、項目レベルでの分析を行っていく予定である。

参考文献

- Arbuckle. J.L. (1999) *Amos 4.0 Programming Reference Guide* SmallWaters Corporation.
池田央(1973). テストII. 東京大学出版会.
豊田秀樹(1998) 共分散構造分析【入門編】朝倉書店