

【原 著】

大学入試センター試験における 科目別得点の非線形因子分析による比較

大津 起夫*

要 約

平成 20 年度から平成 22 年度までの大学入試センター試験（本試験）におけるいくつかの科目の難易度の比較を行った。分析の対象としたものは、大学入試センターが得点調整の対象としている地歴、公民、および理科の科目である。

試験の難度の比較のために、欠測値に対応した非線形因子分析（大津，2004）を用いた。古典的な因子分析のモデルは、潜在変数と観測変数との線形の関係を表す。因子分析は心理テストの分析に広く用いられてきたが、大規模な学力試験の分析を行うためには、非線形関係を表現するモデルを用いる必要がある。

ここでは、観測変数の条件付き平均を潜在変数の区分多項式によって表す非線形因子分析の手法を用いてセンター試験の科目得点を分析した。潜在変数の離散近似を用いることにより、MAR（Missing At Random）と呼ばれる欠測条件への対応を容易に行うことができる。

分析結果を検討すると、科目得点の周辺分布のみによっては、科目間の難しさの違いを十分に検知することができない。適切な得点の調整を行うためには、得点の同時分布の全体に注意を払う必要があることが示唆された。

キーワード：センター試験，難易度，得点調整，非線形因子分析，MAR

1 背景と目的

大学入試センター試験の利用において、受験者がある学習分野から排他的に受験科目を選択し、それらの得点が選抜において共通の尺度を持つものとみなされて利用されることが多い。たとえば、センター試験の「公民」分野では「現代社会」、「倫理」、「政治・経済」の試験が同一の時間枠内で実施されるため、受験者は、これらのうちから排他的に1つの科目を選択しなければならない。選抜を実施する大学においては、選抜資料として重みづけられたセンター試験の科目得点を用いることが多く、この際これらの科目の得点は互いに共通の

尺度による数値であるものとして取り扱われている。しかしながら、年度によっては科目間の得点平均が大きく異なることが時折生じてきた。1979年における第1回の共通第一次学力試験より、これまでに、教科間の得点の違いに対処するために、「得点調整」（事後的に受験者の科目得点を素点から変換すること）が、二度にわたり大学入試センターによって実施されている。現在では、科目得点の周辺分布を比較し、素点の平均が同一教科の得点の科目間で20点以上の差があり、しかもその差が試験問題の難しさの違いによると判断される場合に、平均点差が最大15点となるよう得点調整を行うこととしている（真弓・村上・白旗・吉村・前川，1999）。

* 大学入試センター研究開発部試験評価解析研究部門，JST/CREST
2010年12月20日 受理

しかしながら、異なる科目の試験得点を同一の基準で比較することには、多くの困難がある。厳密に考えるならば、異なる科目の知識を問うことは比較可能とはいえず、同一の尺度によって数値化することはそもそも不可能な試みであると主張することもできる。確かにこのような立場をとることもできるが、現実には多くの入学試験で複数の科目得点を、あたかも同等に比較可能な数値であるかのように扱っている。このような取り扱いが妥当であるとみなすためには、いくつかの仮定をおく必要がある。主要な見方としては次の2つの立場が考えられるだろう。

一つには、たとえ科目が異なっても、それらの成績に共通に反映される特定の能力があり、それを複数の科目得点が幾分なりとも共通に反映しているとする立場である。実際に、センター試験の科目得点の相関を見ると、異なった科目間にもかなり明確な相関があり、複数の科目に共通して測定されている何かしらの能力が存在していると想定することは不自然ではない。ただし、この立場は、必ずしも心理学における一般知能論と同等の主張を行うことにはならない。一般知能論の主張するところは、人間の認知能力の広い範囲にわたって共通に基盤となる能力があり、このパフォーマンスが個人の知的・社会的な適性に重要な影響を与えているというものである。ここでの主張は、学力試験における複数の科目に共通して測定される能力が存在するという事に限定されており、学力試験のパフォーマンスが人間の認知能力を全面的に予測することを主張するものではないし、またそのような仮定を置く必要があるわけでもない^{注1)}。

もう一つの立場は、異なる科目の試験によって測定される能力に共通の要素が含まれておらずとも、それらは互いに補完し、ある一定の比率で交換可能な価値を持つとみなすものである。異なる科目得点によって測定される共通の能力が存在しなくとも、能力を評価する側においてそれらの能力の価値が交換可能なものであるならば、あたかも市場での取引のように一定のレートでの交換が成立するとみなせる。試験における評価が、このような能力の交換価値を示していると考えられることも一つの立場として可能である。伝統的な計量心理学の研究は、主に複数の試験が共通の能力を測

定していることを前提として比較を行おうとしてきたため、このような立場についての研究はほとんど行われていない。しかし、後者の見解は、前者より人間の能力の構造やメカニズムについて要求する仮定が少ないため、前者よりも主張の擁護がより容易である。

伝統的に、心理学における能力測定にかかわる測定技術論(狭義の計量心理学, psychometrics)においては、前者の立場をとることが多い。一方、後者の互換性についての検討材料は、試験の得点データをいくら詳細に分析しても得られず、むしろ能力の需要側(入学試験ならば受け入れ側の学校であり、就職試験ならば企業や政府)についての調査が必要になる。

以下での分析は、基本的には前者の立場をとる。これは、試験によって得られたデータの分析に基づいた統計的な推論によって有意な結論を得ようとする、こちらの視点を取らざるを得ないためである。しかしながら、現実の入学試験の運用を行っている関係者が、複数科目間の得点の比較についてこのような視点に立っていると保証はない。試験運用にかかわる多くの関係者の意識を明らかにする資料は見当たらないが、明白には語られないものの、後者の立場を暗黙の前提としていることも考えられる。このため、以下に示すような推論やこれに類似した方法のみが、科目別得点の同等性について検討する方法ではないことに留意する必要がある。

2 方 法

2.1 分析の枠組み

上に述べた目的に従って、科目別得点を共通の基準によって比較するために、欠測値に対応した非線形因子分析(大津, 2004; 大津・橋本・荘島・石塚, 2006)を用いた分析を、平成20年から平成22年までのセンター試験の科目得点を対象として行う。これらの分析には、研究開発部が保持している個人情報秘匿した「志願者ファイル」(橋本・大津・石岡, 2009a, 2009b, 2011)を用いた。

平成20年度から平成22年度の大学入試センター試験における得点調整対象科目は3つの科目セット(第1セットは「世界史B」, 「日本史B」, 「地理

B], 第2セットは「現代社会」, 「倫理」, 「政治・経済」, 第3セットは「物理 I」, 「地学 I」, 「化学 I」, 「生物 I」からなる。これらの各々について, 本試験における科目間の難易度の比較を行う。各科目の受験者は同一ではないため, 「国語」, 「数学 I・数学 A」, 「数学 II・数学 B」, および「英語」の科目得点をアンカー科目として利用し, 難易度の推定を試みる。以下では, 各年について, 「国語」, 「歴史・地理」, 「公民」, 「英語」の科目得点を対象とする分析(文系科目の分析と呼ぶ)と, 「数学 I・数学 A」, 「数学 II・数学 B」, 「理科」, および「英語」を対象とする分析(理系科目の分析と呼ぶ)を, それぞれ行う。

平成 18 年および平成 19 年の科目得点については, 既に大津・林(2007)にて, また平成 20 年と平成 21 年については大津(2009)にて類似の分析結果を研究開発部の内部資料として報告済みである。ここでは平成 20 年以降について, 後述の「補正済み平均」の推定法の改善を行い, これに基づいて推定値を再計算し, 加えて平成 23 年度の科目得点の新たな分析結果を示す。

2.2 分析のためのモデル

因子分析は, 観測変数が複数の変数に共通する潜在変数と各変数に固有の独自因子との和によって表現されると仮定してモデル推定を行う。ここでは観測変数の条件付平均(潜在変数の値を固定した場合の観測変数の平均)が, 区分多項式(スプライン関数)(de Boor, 2001)によって表されるモデルを用いる。この際, 独自因子は潜在変数の値によらず, 一定の分散に従うと仮定する。しかしながら, 試験の得点を分析する場合には, 満点付近での独自因子による分散は, より平均的な得点における場合よりも小さいと思われる。そこでそれぞれの変数(素点) X_j を次の式(逆正弦変換)によって Y_j に変換し, 因子分析モデルの適用がより妥当となる変数を構成する。

$$Y_j = f_j(X_j) = \sin^{-1} \sqrt{\frac{X_j}{X_j \text{の満点}}} \quad (1)$$

大津(2004)による非線形の因子分析モデルは次のようなものである。

$$Y_j = \eta_j(\theta) + \varepsilon_j \quad (2)$$

ここで θ は標準正規分布に従う直接には観測されることのない確率変数(潜在変数)であり, ここでは各科目に共通な能力を表すものと仮定する。また ε_j は θ および他の ε_k , ($k \neq j$) と互いに独立な確率変数(独自因子)であり, それぞれ独立に分散 ψ_j 平均ゼロの正規分布に従うものとする。また η_j はそれぞれの変数に対応した区分多項式(スプライン)であるとする。

ここで潜在変数の次元は 1 次元とし, 正規分布を 50 個の等分位区間に分割し, それぞれの区間を平均 θ_k , ($k = 1, 50$) によって代表させることにより, 離散的に近似した。また各 η_j は, 離散化された θ の最小値 θ_1 , 最大値 θ_{50} および $\theta = 0$ (中位数)の 3 点を節点とする 2 次多項式による B スプラインによって推定した。さらに, 分析対象者のすべてが科目 j を受験したら得られるであろう科目の平均点の推定値を次の式で求める。

$$E_{\theta, \varepsilon_j}[g_j(Y_j) = f_j^{-1}(\min(\max(0, Y_j), \pi/2))] \quad (3)$$

ここでは θ は上述のように 50 個の離散値で近似するが, さらにそれぞれの ε_j も同様に 50 個の離散値で近似して推定を行う。このようにして得られた各科目の平均点の推定値を「補正済み平均」と呼ぶことにする^{注2)}。

また, 「共通性の推定値」とは非線形の因子分析モデルによって推定された η_j と ψ_j を用いて定義される次の値を示す。

$$\hat{C}_j = 1 - \frac{\hat{\psi}_j}{\text{Var}_{\theta}(\hat{\eta}_j(\theta)) + \hat{\psi}_j} \quad (4)$$

「共通性」とは線形の因子分析モデルで用いられる用語であるが, 上式はそれにほぼ対応する指標であるため, 線形モデルの場合に準じて「共通性」と呼ぶことにする。

2.3 欠測への対応

センター試験においては, 全ての科目得点が観測されるのではなく, 各受験者については一部の科目得点しか観測されない。ここで, 欠測メカニズムに MAR (Missing At Random) (Rubin, 1976) と呼ばれる仮定をおき, その上でモデルの推定を行う。MAR の仮定は, 欠測の確率が観測されている値のみに依存し, 欠測している変数の(もし受験したら得られるであろう)値には依存しないこと

を求める。試験における科目選択において MAR の仮定が厳格に成立していることは現実には疑わしく、またこの仮定からの違反の程度を正しく推定することも難しい。しかしながら、MAR の仮定の下での推定は、欠測メカニズムを無視したナイーブな推定^{注3)}よりは妥当な結果を導く。より精密な欠測メカニズムを仮定することも可能ではあるが、これを指定するためには、実際に特定の科目を受験しなかった学生がその科目を受験した場合に得られるであろう得点についての情報が必要になり、センター試験の成績得点だけからの分析では推論が困難である。これらの制約を踏まえ、ここでは MAR の仮定のもとでの推定を行う。

実際、理科の科目得点を詳細に検討すると、明らかに MAR の仮定に違反すると推測される特徴が一部に存在するため、ここでの推定の信頼度には限界がある。またさらに注意すべき点として、国語における「古典」「漢文」の選択がある。現在のセンター試験の「国語」の問題構成は、現代文(近代以降の文章)についての大問が2題と、古文が1題、漢文が1題のあわせて4題構成であり、各々50点の配点となっている。一部の大学においては、国語の成績として、これらの一部(おもに現代文の成績)のみを要求し、古文および漢文の得点を要求しない場合がある。解答データを検討すると国語受験者のうち、6%から7%程度において古文の各項目の解答欄が空白であり、また漢文では12%から13%程度において各項目の解答欄が空白である。この問題に対応するために、現代文の得点のみを分析に利用することも考えられるが、大問別の識別力を検討すると、現代文の問題は大きくないため、「古文」「漢文」の大問を利用したほうが望ましいと判断した。また、分析が過度に複雑になることを避けるため、これらの空白の場合の得点を特に欠測とみなす処理は行っていない。このため他の科目得点との同時分布の分析においては、「国語」の科目得点の識別力の推定に偏りが生じている可能性がある。さらに、「数学II・数学B」においても、選択問題があるため、厳密にはこれら大問ごとの得点を分析するほうが望ましいが、ここでの主な関心が、「歴史・地理」「公民」「理科」の科目間の難易度比較にあるため、科目得点のみを分析の対象とした。また、「英語」については筆記試験とリスニングの得点の合計点を

分析対象とした。

3 平成20年センター試験本試験の分析

3.1 分析1：平成20年文系科目の分析

平成20年(2008年)のセンター試験の本試験において、「国語」と「英語」(筆記試験およびリスニング)の2科目(3枚の解答)ともに受験記録があるもの471,515名を分析対象とする。これらの受験者の科目得点「国語」、「世界史A」、「世界史B」、「日本史A」、「日本史B」、「地理A」、「地理B」、「現代社会」、「倫理」、「政治・経済」、「英語」および「英語リスニング」を分析に用いる。「英語」と「英語リスニング」については、得点を合計し一つの変数とする。他の科目はそれぞれを1つの変数として扱い、合計11変数での分析を行う。

さらに、図1には、「世界史B」、「日本史B」、および「地理B」について、 $\eta_j(\theta)$ を逆変換して、素点スケールに戻したもの、つまり $g_j(\eta_j(\theta))$ の形状を示す。また「現代社会」、「倫理」、「政治・経済」についての同様の分析結果を図2に示す。地歴科目においては、「世界史B」がやや難しく、他の2科目はほぼ同等の難度であると推測される。また公民の3科目は、すべての学力層において、ほぼ同等の難度と推定される。

表1の分析においては、各 η_j は $\theta=0$ の箇所にも1つの内部節点をもつ区分2次多項式としているが、 η_j をより複雑な関数に置き換えても、推定結果の変化量は大きくはない。表2は、スプラインの内部節点を $(-0.674, 0.0, 0.674)$ の3点(それぞれ標準正規分布の1/4, 1/2, 3/4分位点)として推定を行った場合の結果を示した。表中の「条件付き平均の差」は、表1の推定(内部節点を $\theta=0.0$ の1箇所とした場合)における $g(\eta_j^{M1}(\theta_k))$ から、内部節点を3箇所として推定した値 $g(\eta_j^{M3}(\theta_k))$ を引いた差の $k=1, \dots, 50$ における最小値と最大値を示したものである。

結果を検討すると「世界史A」、「日本史A」および「英語」を除いては、100点満点に換算して最大値・最小値のいずれも-1点から1点以内の範囲にあり、ほとんど推定結果が変わらない。このうち「英語」については違いの最大値が4.8となっているが、これは θ の下端 θ_1 において生じて

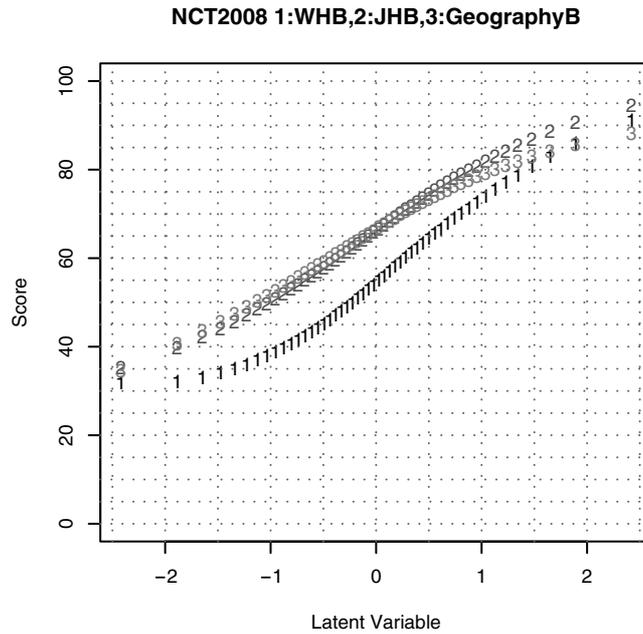


図 1 平成 20 年 1:「世界史 B」, 2:「日本史 B」, 3:「地理 B」得点の条件付平均の推定値 (横軸は潜在変数, 縦軸は科目得点)

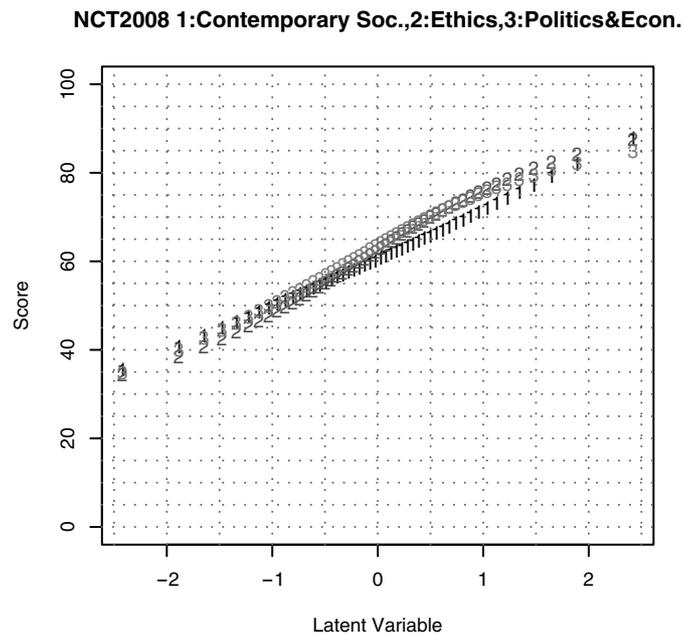


図 2 平成 20 年 1:「現代社会」, 2:「倫理」, 3:「政治・経済」得点の条件付平均の推定値 (横軸は潜在変数, 縦軸は科目得点)

おり, これを除くと違いの最大値は 250 点中 1.87 点, 最小値は -2.01 点である.

モデルの内部節点の違いによって, 表 2 のような結果が得られることから, 受験者数が少なく欠測者の科目得点についての η_j の推定が幾分変動しやすいことがわかる. また, 受験者数が多く欠測

の存在しない場合についても, θ の端点において η_j の形状が不安定となることがあることがわかる.

受験者数の少ない科目において η_j の推定値の 2 つのモデルにおける違いが 3 点ほどになる箇所があるが, 独自因子を加えた $g_j(Y_j) = g_j(\eta_j(\theta_k) + \varepsilon_j)$ の分布関数は, 独自因子によって平滑化されるた

表 1 分析1 平成 20 年文系科目の特徴

科目	国語	世界史 A	世界史 B	日本史 A	日本史 B	地理 A	地理 B
満点	200	100	100	100	100	100	100
本試験受験者	481,315	2,164	93,928	4,260	143,676	5,811	107,519
分析対象件数	471,515	1,997	92,433	3,962	141,077	5,554	106,598
本試験平均	121.64	49.28	58.98	55.95	64.27	56.83	66.36
分析対象平均	122.38	49.99	59.19	56.47	64.53	57.09	66.48
補正済み平均	122.45	55.42	55.57	63.35	65.30	61.05	65.07
共通性の推定値	0.563	0.588	0.527	0.478	0.551	0.405	0.534

科目	現代社会	倫理	政治・経済	英語 (+リスニング)
満点	100	100	100	250
本試験受験者	174,686	51,134	80,598	490,853 (リス)
分析対象件数	172,772	50,570	79,029	471,515
本試験平均	60.55	67.58	63.73	(125.26 + 29.45)
分析対象平均	60.67	67.70	63.95	156.38
補正済み平均	60.47	61.93	62.63	156.60
共通性の推定値	0.519	0.588	0.536	0.730

本試験平均における英語（筆記）とリスニングの値は、それぞれの受験者集団についてのものである。英語（筆記）の本試験受験者数は 497,101 名であり、リスニングより 6,248 名多い。ただし、リスニングのみ受験し、筆記試験を欠席しているものも若干名存在する。

表 2 分析1 3つの内部節点を持つモデルによる推定結果

科目	国語	世界史 A	世界史 B	日本史 A	日本史 B	地理 A	地理 B
補正済み平均	122.35	55.44	55.51	63.27	65.25	61.04	65.01
共通性の推定値	0.563	0.583	0.526	0.478	0.551	0.407	0.535
条件付き平均の差 (最大)	0.89	2.09	0.64	2.31	0.48	0.40	0.41
(最小)	-1.61	-2.46	-0.29	-3.03	-0.68	-0.44	-0.83

科目	現代社会	倫理	政治・経済	英語 (+リスニング)
補正済み平均	60.44	61.89	62.59	156.42
共通性の推定値	0.519	0.588	0.537	0.732
条件付き平均の差 (最大)	0.25	0.07	0.45	4.80
(最小)	-0.40	-0.15	-0.82	-2.01

め、差が相殺されて小さくなる傾向がある。潜在変数の値 θ_k , ($k = 1, \dots, 50$) の各々について独自因子 ε_j を 50 個の離散点によって近似することにより Y_j の分布を 2500 個の離散点によって近似することができる。これらの点を昇順に並べることにより、 $g(Y_j)$ の周辺分布の 2500 個の分位点の推定値が得られる。これらの分位点の違いは、推定された $g(Y_j)$ の周辺分布の違いを示す。2つの推定結果による分位点の違いを各科目について検討し、差の平均、標準偏差、最大値および最小値を表 3 に示した。いずれの科目についても、独自因子を加えた周辺分布の違いはほとんど無視し得る

ものであることが分かる。「英語」では幾分大き目の違いが生じているが、最大の違いは $g_j(Y_j)$ の最小値の箇所で起きており、これ以外の箇所で違いの最大値は 1.61 である。

この分析例においては、データ件数が多いためモデルパラメータ数増加による対数尤度の増加はかなり大きく（約 390）、パラメータ数を増加させることによる利得は存在するが、この変更によって解釈上の大きな変更は生じないと判断される。また他の分析例への適用においては、内部節点を増加させることにより η_j の単調性がいくつかの科目において損なわれる場合もみられた。複数のケー

表 3 分析1 内部節点数の変更に伴う分位点の変化量

科目	国語	世界史 A	世界史 B	日本史 A	日本史 B	地理 A	地理 B
平均	0.09	-0.02	0.06	0.09	0.06	0.01	0.06
標準偏差	0.14	0.28	0.07	0.11	0.07	0.06	0.07
最大値	0.60	0.88	0.31	0.56	0.24	0.26	0.31
最小値	-1.08	-0.49	-0.21	-0.55	-0.48	-0.23	-0.72

科目	現代社会	倫理	政治・経済	英語 (+リスニング)
平均	0.04	0.04	0.04	0.18
標準偏差	0.06	0.03	0.06	0.53
最大値	0.29	0.11	0.25	3.22
最小値	-0.38	-0.15	-0.30	-1.09

2500 個の値 $g_j(\eta_j(\theta_k) + \varepsilon_i)$, ($k = 1, \dots, 50$; $i = 1, \dots, 50$) を昇順に整列して 2500 個の分位点を求め、0 点と満点を超えるものは範囲内に変更する。表 1 の結果から得られる分位点の値から、表 2 の結果に基づく分位点を値を引いた差について平均、標準偏差、最大値および最小値を示した。

スについて同一のモデルによる推定を行うことを考慮し、以降では内部節点が 1 個のモデルによって分析を行う。

3.2 分析 2：平成 20 年理系科目の分析

平成 20 年 (2008 年) のセンター試験において、「数学 I・数学 A」および「英語」(筆記試験およびリスニング) の 2 科目 (3 枚の解答) を本試験で受験したものの 346,172 名を分析対象とする。これらの受験者の科目得点「数学 I・数学 A」, 「数学 II・数学 B」, 「物理 I」, 「地学 I」, 「理科総合 A」, 「化学 I」, 「理科総合 B」, 「生物 I」, 「英語」および「英語リスニング」を分析に用いる。数学②の時間に選択可能な「数学 II」, 「工業数理基礎」, 「簿記・会計」, 「情報関係基礎」の得点については、受験者数が比較的少数であり、推定に困難を生じるためここでの分析の対象とはしない。ただし、これらの科目の受験者も「数学 I・数学 A」と「英語」(リスニングを含む) をともに受験しているならば、分析対象に含む。分析 1 と同様に「英語」および「英語リスニング」については、得点を合計しひとつの変数とする。他の科目はそれぞれの科目得点を変数とし、合計 9 個の変数を分析に用いる。

モデル推定の設定は、分析対象となる変数の個数をのぞいては分析 1 と同様とする。50 個の潜在的な離散分布を用い、潜在変数の最小値、最大値、および $\theta = 0$ (中位数) の 3 点を節点とする 2 次多項式のスプライン関数を利用する。

表 4 に、それぞれの変数の特徴をあらわす統計

量の一覧を示す。さらに、図 3 には、得点調整の対象となっている理科 4 科目について、潜在変数を固定して得られる科目得点と条件付平均の推定値 $\eta_j(\theta)$ を逆変換して、素点スケールに戻したものの、つまり $g_j(\eta_j(\theta))$ の形状を分析 1 と同様に示す。「物理 I」と「化学 I」の曲線が極めて類似しており、これら 2 つの教科はほぼ同等の難度を持っていたと推測される。

文系科目の分析 (分析 1) においては、共通性の推定値は「地理 A」と「日本史 A」とが 0.4 台で幾分小さく、「英語」が 0.730 で大きい、他の科目はいずれも 0.5 台であり、大きな違いがない。一方、分析 2 では、数学の 2 つの科目および「物理 I」, 「化学 I」で共通性が大きく、「生物 I」と「英語」で中間的であり、「地学 I」, 「理科総合 A」, 「理科総合 B」で小さい。特に「理科総合 A」および「理科総合 B」の共通性の推定値は 0.2 台である。これらの科目得点が、分析に用いた他の科目と共通の能力を良く表していると思えるのは、少なくともこの分析からは難しい。また「地学 I」の共通性の推定値は 0.353 であり、これも他の科目の得点による予測が難しいことがわかる。「地学 I」は現状での得点調整の対象となっている科目であるが、平成 21 年度の分析にて後述するように、得点の挙動に注意すべき点があり、MAR の仮定からの逸脱が大きいことが疑われる。

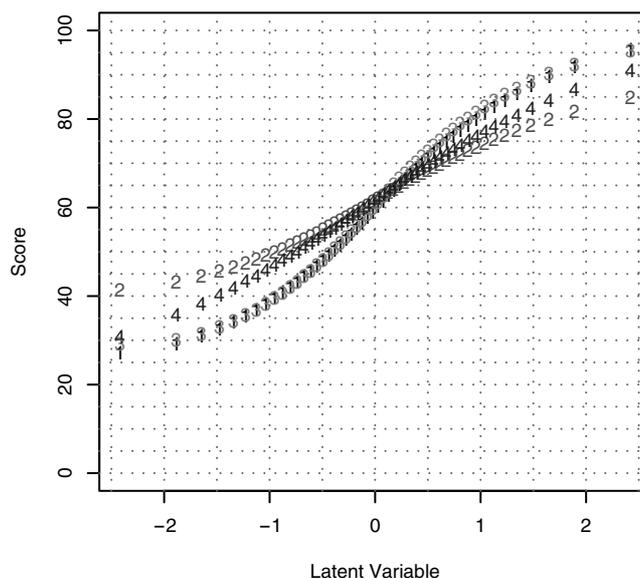
また、「数学 I・数学 A」の補正済み平均は、分析対象となった全ての受験者が受験しているため、分析対象平均と本来等しくなるべきであるが、実

表 4 分析2 平成 20 年理系科目の特徴

科目	数学 IA	数学 IIB	物理 I	地学 I	理科総合 A	化学 I	理科総合 B
満点	100	100	100	100	100	100	100
本試験受験者	350,198	317,103	142,233	26,841	33,472	199,951	17,614
分析対象件数	346,172	312,473	139,749	24,865	30,581	195,710	15,118
本試験平均	66.31	51.01	64.55	59.68	48.00	64.21	61.31
分析対象平均	66.53	51.26	64.89	60.18	48.07	64.58	61.95
補正済み平均	66.99	49.94	59.46	61.32	51.47	60.27	63.41
共通性の推定値	0.748	0.832	0.653	0.353	0.253	0.699	0.201

科目	生物 I	英語 (+リスニング)
満点	100	250
本試験受験者	176,766	490,853 (リス)
分析対象件数	162,886	346,172
本試験平均	57.64	(125.26 + 29.45)
分析対象平均	58.61	161.91
補正済み平均	61.13	162.55
共通性の推定値	0.534	0.520

NCT2008 1:PhysI,2:EarthI,3:ChemI,4:Biol

図 3 平成 20 年 1:「物理 I」, 2:「地学 I」, 3:「化学 I」, 4:「生物 I」
得点の条件付平均の推定値 (横軸は潜在変数, 縦軸は科目得点)

際にはわずかであるが過大評価 (0.46 点) となっている。ただし、周辺分布において特に大きな食い違いの見られる部分は存在しない。

また「英語」も同様に、分析対象平均と補正済み平均は同一になるべきであるが、250 点中 0.64 点の過大評価となっており分析 1 より差が大きい。周辺分布を検討すると、極端に低い得点層において、予測分布と実際のデータの違いが大きいこと

がわかる。

4 平成 21 年センター試験本試験の分析

4.1 分析 3: 平成 21 年文系科目の分析

平成 21 年についても、平成 20 年文系科目 (分析 1) と同様の方法で分析を行う。分析対象は、「国

表 5 分析3 平成 21 年文系科目の特徴

科目	国語	世界史 A	世界史 B	日本史 A	日本史 B	地理 A	地理 B
満点	200	100	100	100	100	100	100
本試験受験者	484,871	2,187	94,106	4,365	144,327	5,501	109,616
分析対象件数	475,341	2,031	92,744	4,085	141,553	5,270	108,740
本試験平均	115.46	44.18	62.70	46.51	57.94	54.70	64.45
分析対象平均	116.17	44.40	62.87	46.97	58.17	54.83	64.56
補正済み平均	116.15	51.18	59.02	55.01	59.02	58.57	63.46
共通性の推定値	0.534	0.668	0.558	0.584	0.583	0.385	0.547

科目	現代社会	倫理	政治・経済	英語 (+リスニング)
満点		100	100	250
本試験受験者	169,711	53,116	82,804	494,342 (リス)
分析対象件数	167,908	52,522	81,198	475,341
本試験平均	60.19	71.51	69.31	(115.02 + 24.03)
分析対象平均	60.31	71.63	69.55	140.61
補正済み平均	60.36	65.70	67.95	140.68
共通性の推定値	0.539	0.624	0.592	0.703

本試験平均における英語（筆記）とリスニングの科目得点の平均は、それぞれの受験者についてのものである。英語（筆記）の本試験受験者数は 500,297 名であり、リスニングより 5,955 名多い（リスニングのみ受験し、英語筆記試験を欠席した者も若干名存在する）。

NCT2009 1:WHB,2:JHB,3:GeographyB

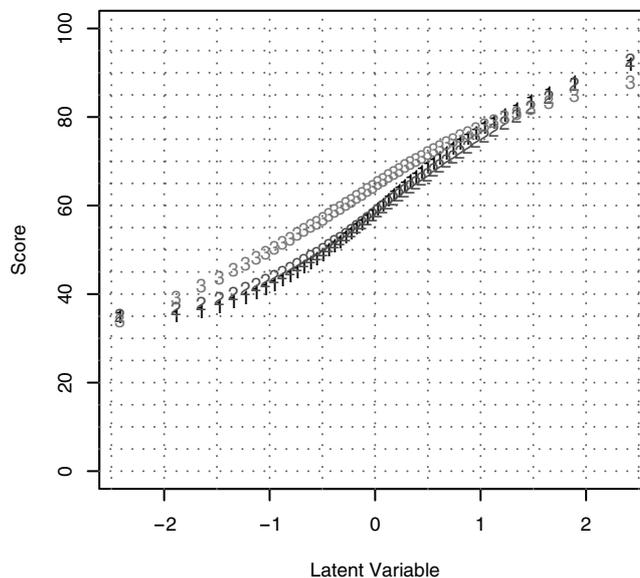


図 4 平成 21 年 1:「世界史 B」, 2:「日本史 B」, 3:「地理 B」得点の条件付平均の推定値 (横軸は潜在変数, 縦軸は科目得点)

語」および「英語」と「英語リスニング」をともに本試験で受験した 475,341 名である。

表 5 に、それぞれの変数の特徴をあらわす統計量の一覧を示した。また図 4 には、得点調整の対象である地歴科目「世界史 B」, 「日本史 B」, 「地

理 B」の条件付期待値の推定値 $g_j(\eta_j(\theta))$ のグラフを示した。さらに、図 5 には、得点調整の対象である公民科目「現代社会」, 「倫理」, 「政治・経済」の 3 科目について、同様のものを示した。

共通性の相対的な大小関係は、平成 20 年（分析

NCT2009 1:Contemporary Soc.,2:Ethics,3:Politics&Econ.

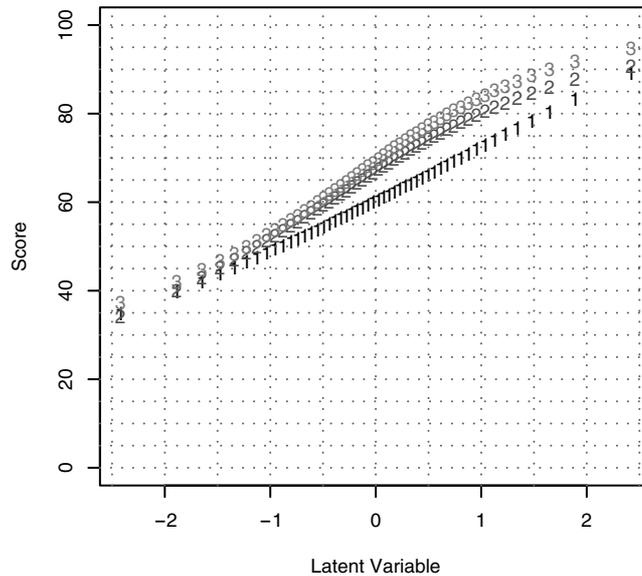


図 5 平成 21 年 1:「現代社会」, 2:「倫理」, 3:「政治・経済」得点の条件付平均の推定値 (横軸は潜在変数, 縦軸は科目得点)

表 6 分析4 平成 21 年理系科目の特徴

科目	数学 IA	数学 IIB	物理 I	地学 I	理科総合 A	化学 I	理科総合 B
満点	100	100	100	100	100	100	100
本試験受験者	354,609	319,045	143,646	25,921	30,427	200,411	17,175
分析対象件数	350,579	314,971	141,493	24,439	28,390	196,854	14,986
本試験平均	63.96	50.86	63.55	51.85	56.59	69.54	58.35
分析対象平均	64.16	51.03	63.84	52.19	56.65	69.86	58.88
補正済み平均	64.39	49.65	58.03	53.13	60.24	65.41	60.88
共通性の推定値	0.733	0.795	0.722	0.408	0.272	0.723	0.230

科目	生物 I	英語 (+リスニング)
満点	100	250
本試験受験者	176,043	494,342 (リス)
分析対象件数	164,224	350,579
本試験平均	55.85	(115.02+24.03)
分析対象平均	56.68	145.52
補正済み平均	59.13	145.83
共通性の推定値	0.528	0.531

1) とかなり共通した傾向が見られる。この分析の枠組みでは、平成 20 年、平成 21 年ともに「英語」の共通性が大きく、ついで「世界史 A」と「倫理」で大きい。ただし、「日本史 A」の共通性は平成 20 年(分析 1)においては 0.478 で「地理 A」について小さいが、平成 21 年度では 0.58 を超えている。

4.2 分析 4: 平成 21 年理系科目の分析

平成 21 年の理系科目についても、平成 20 年(分析 2)と同様の科目を対象として分析を行う。分析 2 と同じく、「数学 I・数学 A」および「英語」(筆記試験およびリスニング)の 2 科目ともに本試験にて受験したものの 350,579 名を分析対象とする。分析結果を表 6 に示し、得点調整対象である 4 科

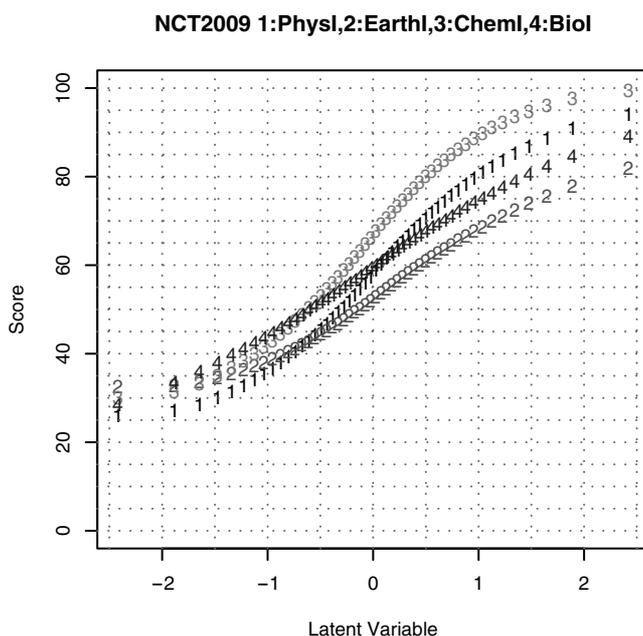


図 6 平成 21 年 1:「物理 I」、2:「地学 I」、3:「化学 I」、4:「生物 I」
得点の条件付平均の推定値

目の条件付平均の推定値 $g_j(\eta_j(\theta))$ のグラフを図 6 に示す。

表 6 において、共通性の推定値の大小は、分析 2 と同様の傾向が見られる。「数学 I・数学 A」、 「数学 II・数学 B」、 「物理 I」 および 「化学 I」 の共通性が大きく、一方「理科総合 A」と「理科総合 B」の共通性が小さく推定されている。また、これら 2 科目について「地学 I」の共通性が小さい。

平成 21 年度においては、「化学 I」と「地学 I」の得点平均に 18 点に近い差が生じたが、補正済み平均の差は約 12.28 点であり、欠測構造を考慮すると平均得点の差はかなり小さくなる。ただし、「化学 I」の科目得点が、得点調整対象となっている他の科目得点より高い傾向は変わらない。また、補正済み平均の差は 13 点を越えないものの、図 6 から、学力階層別に検討すると上位層 ($\theta = 1 \sim 2$ あたり) においては「地学 I」と「化学 I」間にかなり大きな得点差が生じていることがわかる^{注 4)}。また平成 20 年の理系科目の分析結果 (分析 2) と比較すると、特に「化学 I」の得点が高かった (試験が易しかった) ことが、科目間の得点差を広げることには大きな影響を与えていることがわかる。

4.3 分析 5:平成 21 年「地学 I」科目得点の特徴

前項の分析 4 において、平成 21 年の理系科目の科目得点について検討し、特に高学力層において「化学 I」と「地学 I」の得点差が大きい傾向にあることが示唆された。ただし「地学 I」の得点の様相には、かなり特異な点があり、これら注意すべき特徴が、前項の分析からだけでは十分にはわからない。これらの特徴は、分析 4 および分析 2 の結果を検討するうえで留意すべき点である。

「地学 I」得点の分布において特異な点は、十分な準備をせずに「地学 I」を選択し解答していると疑われる受験者が相当数存在することである。図 7 は、平成 21 年大学入試センター試験の「地学 I」と他の理科科目得点との同時分布 (分析対象は分析 4 と同様) を示したものである。それぞれの円の面積が該当する箇所相対的な頻度を示している。いずれも横軸が「地学 I」の科目得点を示し、縦軸が他の理科の科目得点を示す (「物理 I」は「地学 I」と同じ時間枠で実施されるため、「地学 I」との同時受験者がいない)。また、表 7 に分析対象者のうちの同時受験者についての、理科科目得点間の相関係数を示す。

これらを検討すると、「理科総合 B」を除き、「地

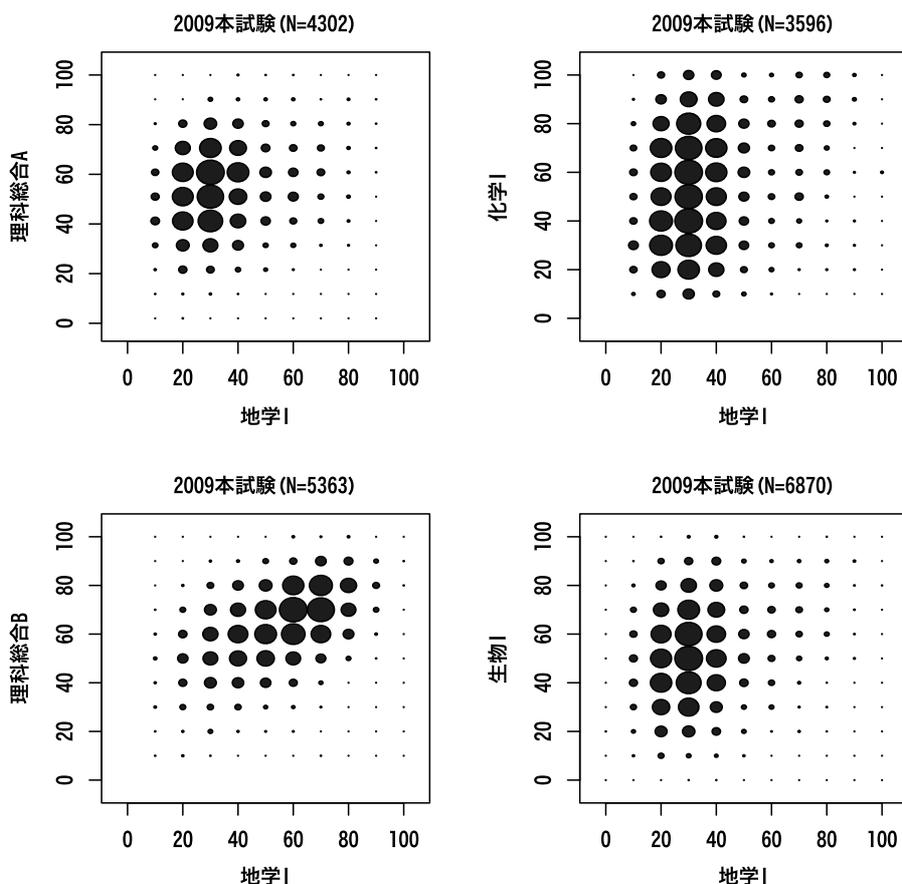


図 7 平成 21 年「地学 I」と他の理科科目得点の同時分布

表 7 平成 21 年「地学 I」と理科他科目得点との相関係数 (分析対象者)

	物理 I	地学 I	理総 A	化学 I	理総 B	生物 I
物理 I	1.000	NA	0.504	0.733	0.328	0.394
地学 I	0	1.000	0.171	0.212	0.513	0.212
理総 A	3,920	4,302	1.000	NA	0.445	0.404
化学 I	123,797	3,596	0	1.000	0.304	0.610
理総 B	6,314	5,363	2,592	6,428	1.000	NA
生物 I	13,659	6,870	22,014	65,774	0	1.000
分析対象者数	141,493	24,439	28,390	196,854	14,986	164,224

「数学 I・数学 A」, 「英語」および「英語リスニング」の同時受験者のうち対象となる 2 科目を同時受験した者の数と, それらの受験者における 2 科目間の相関を示す. 上三角部分が相関係数. 下三角部分は同時受験者数を示す.

学 I」の科目得点と他の理科の科目得点との関係が著しく小さく, また同時受験者の「地学 I」の平均点も低いことが見て取れる. 通常は, 理科の一つの科目と他の科目の間には, 相当の相関があると期待されるが, 「地学 I」と他の理科の科目得点の相関は, 際立って低い.

また, さらにデータを検討すると, このような傾向は, 地域によって程度が異なることがわかる.

表 8 には, 東京と近隣の県における, 分析対象者の理科科目得点間の相関を示す. また「地学 I」と理科の他科目得点との同時分布を図 8 に示す. この地域では, 受験者一人あたりの受験科目数は少ない傾向があり, 「地学 I」と他の理科の同時選択者が相対的に少なく, 「地学 I」の科目得点と「理科総合 B」を除く他の理科科目得点との相関は, より大きい傾向にある.

表 8 平成 21 年「地学 I」と理科他科目得点との相関係数（東京近隣の分析対象者）

	物理 I	地学 I	理総 A	化学 I	理総 B	生物 I
物理 I	1.000	NA	0.471	0.721	0.393	0.462
地学 I	0	1.000	0.249	0.456	0.462	0.373
理総 A	310	207	1.000	NA	0.524	0.420
化学 I	19,936	370	0	1.000	0.293	0.642
理総 B	218	312	193	245	1.000	NA
生物 I	1,547	600	462	10,106	0	1.000
分析対象者数	27,653	4,638	989	36,075	816	21,300

「数学 I・数学 A」, 「英語」および「英語リスニング」の同時受験者のうち対象となる 2 科目を同時受験し, 在学あるいは卒業した高校が埼玉, 千葉, 東京, 神奈川のいずれかの都県にある者の数と, それらの受験者における 2 科目間の相関を示す. 上三角部分が相関係数. 下三角部分は同時受験者数.

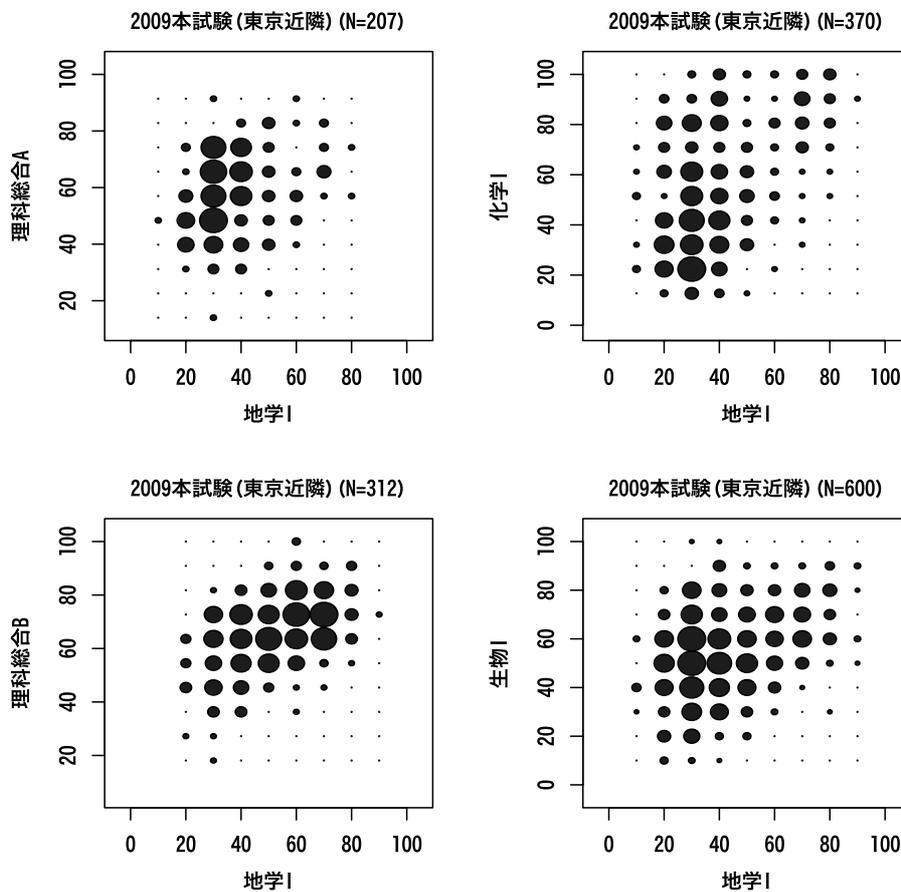


図 8 平成 21 年「地学 I」と他の理科科目得点の同時分布（東京近隣）

4.4 分析 6：平成 21 年理系科目の分析（東京近隣）

このように, 他の科目で高得点でありながら, 「地学 I」で低い得点である受験者が相当数存在すると, 「地学 I」科目得点の学力階層別の難度が過大評価される可能性がある. そこで, 分析 4 の対

象者のうち, このような傾向の薄い傾向が見られる東京近隣 4 都県（埼玉, 千葉, 東京, 神奈川）の高校に在学中か, あるいはそれらの高校を卒業した受験者 65,513 名に対象を限定し, 分析をおこなう.

表 9 に受験者の卒業（見込み）高校の所在地を限定した分析結果を示す. また, 得点調整対象で

表 9 分析6 平成 21 年理系科目の特徴 (千葉, 埼玉, 東京, 神奈川)

科目	数学 IA	数学 IIB	物理 I	地学 I	理科総合 A	化学 I	理科総合 B
満点	100	100	100	100	100	100	100
分析対象件数	65,513	57,278	27,653	4,638	989	36,075	816
分析対象平均	64.57	53.84	64.41	60.66	63.19	71.12	61.89
補正済み平均	64.77	52.43	60.62	57.09	67.76	67.28	62.06
共通性の推定値	0.737	0.799	0.729	0.540	0.378	0.696	0.234

科目	生物 I	英語 (+リスニング)
満点	100	250
分析対象件数	21,300	65,513
分析対象平均	60.38	152.27
補正済み平均	62.06	152.53
共通性の推定値	0.562	0.578

NCT2009(Tokyo neib.) 1:PhysI,2:EarthI,3:ChemI,4:Biol

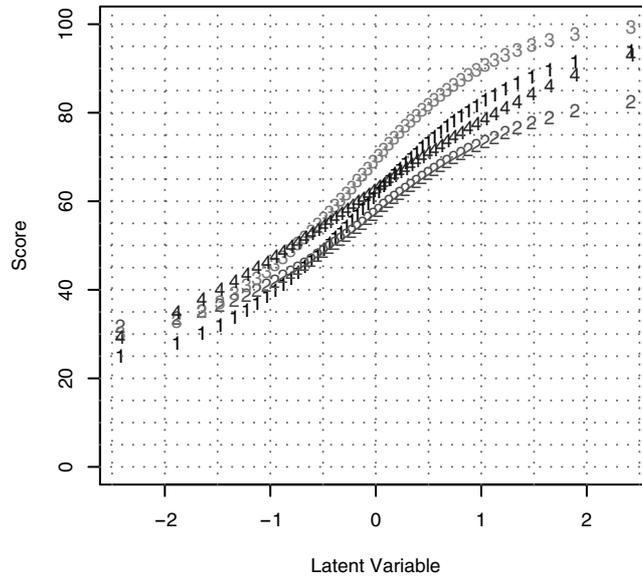


図 9 平成 21 年 (東京近隣) 1:「物理 I」, 2:「地学 I」, 3:「化学 I」, 4:「生物 I」
得点の条件付平均の推定値

ある 4 科目の条件付平均の推定値 $g_j(\eta_j(\theta))$ のグラフを図 9 に示す。この地域における「地学 I」の科目得点の平均は 60.66 であり、地域を限定しない場合 (52.19) に比べて、かなり高い。また補正済み平均は 57.09 であり、観測された科目得点の平均より小さくなっている。これより、この地域においては、分析対象となった科目において地域内で相対的に高学力の志願者が、「地学 I」を選択していると推測される。

「地学 I」と「化学 I」の科目得点の分析対象者の平均得点の差は、この分析においては、10.46 で

あり、地域を限定しない場合 (分析 4) の差 17.67 に比べてかなり小さい。また、補正済み平均における差は、10.18 であり、地域を限定しない場合の差 12.28 より小さい。しかしながら、図 9 を見ると、潜在変数 θ が 1 付近での、これら 2 科目の条件付平均の差はなお大きい。この地域に受験者を限定しても、分析 5 で指摘した傾向が完全になくはしないため、なお受験者の特性による「地学 I」難度の過大評価の傾向は残っていると推測されるが、平成 20 年度の分析 (分析 2) や、大津・林 (2007) における分析結果と比較すると、分析 4

における「地学 I」と「化学 I」の成績上位者における得点差は例年より大きいと思われる。

5 平成 22 年センター試験本試験の分析

5.1 分析 7：平成 22 年文系科目の分析

さらに平成 22 年についても平成 20 年および平成 21 年と同様の方法で分析を行う。分析対象は、「国語」および「英語」と「英語リスニング」をもとに本試験で受験した 487,985 名である。

表 10 に、それぞれの変数の特徴を表す統計量の一覧を示し、また図 10 には、得点調整の対象である地歴科目「世界史 B」,「日本史 B」,「地理 B」の条件付平均の推定値 $g_j(\eta_j(\theta))$ のグラフを示した。さらに、図 11 には、得点調整の対象である公民科目「現代社会」,「倫理」,「政治・経済」の 3 科目について、同様のものを示した。

表 10 における共通性の推定値を検討すると、平成 20 年および平成 21 年の結果と類似した値が求められている。文系科目に共通する潜在変数と各科目得点間の関連の強さが、年度を超えて安定していることが伺える。また、平成 22 年度の分析結果から、文系の得点調整対象科目間で大きな難易度の違いはないことがわかる。実際、補正済み平均の最大の差は「世界史 B」(56.22) と「地理

B」(63.72) 間についてのものであり、7.5 点に留まっている。しかし、より詳細に検討すると、「地理 B」の成績の最上位層においては、「世界史 B」および「日本史 B」より $g_j(\eta(x))$ の推定値が小さくなっており、満点を取りづらい試験であったことがわかる。

5.2 分析 8：平成 22 年理系科目の分析

平成 22 年の理系科目についても、平成 20 年(分析 2)、平成 22 年(分析 4)と同様の科目を対象として分析を行う。同じく、「数学 I・数学 A」および「英語」(筆記試験およびリスニング)の 2 科目ともに本試験にて受験したものの 364,187 名を分析対象とする。分析結果を表 11 に示し、得点調整対象である 4 科目の条件付平均の推定値 $g_j(\eta_j(\theta))$ のグラフを図 12 に示す。

得点調整対象の 4 科目のうち、「物理 I」と「化学 I」とはほぼ同等の難度であると推測されるが、「地学 I」および「生物 I」は得点が高く、特に成績の中位から下位層にかけて違いが大きくなっている。本試験の平均点差は「物理 I」と「生物 I」の間で最大であるが、この点差は補正済み平均において、さらに拡大している。

また、平成 21 年度の理系科目の分析においてみられたのと同様の「地学 I」についての傾向が平成 22 年についてもみられる。理科の科目間について

表 10 分析 7 平成 22 年文系科目の特徴

科目	国語	世界史 A	世界史 B	日本史 A	日本史 B	地理 A	地理 B
満点	200	100	100	100	100	100	100
本試験受験者	497,431	1,979	91,118	4,094	151,792	4,980	110,093
分析対象件数	487,985	1,797	89,734	3,781	149,132	4,748	109,289
本試験平均	107.62	52.31	59.62	48.42	61.51	53.58	65.11
分析対象平均	108.21	53.34	59.82	49.12	61.79	53.91	65.20
補正済み平均	108.16	57.86	56.22	58.18	62.62	57.72	63.72
共通性の推定値	0.524	0.637	0.558	0.527	0.573	0.351	0.563

科目	現代社会	倫理	政治・経済	英語 (+リスニング)
満点	100	100	100	250
本試験受験者	171,419	55,849	89,887	506,898 (リス)
分析対象件数	169,615	55,207	88,295	487,985
本試験平均	58.76	68.66	59.16	(118.14 + 29.39)
分析対象平均	58.90	68.79	59.41	149.23
補正済み平均	59.00	62.94	57.73	149.15
共通性の推定値	0.568	0.646	0.576	0.728

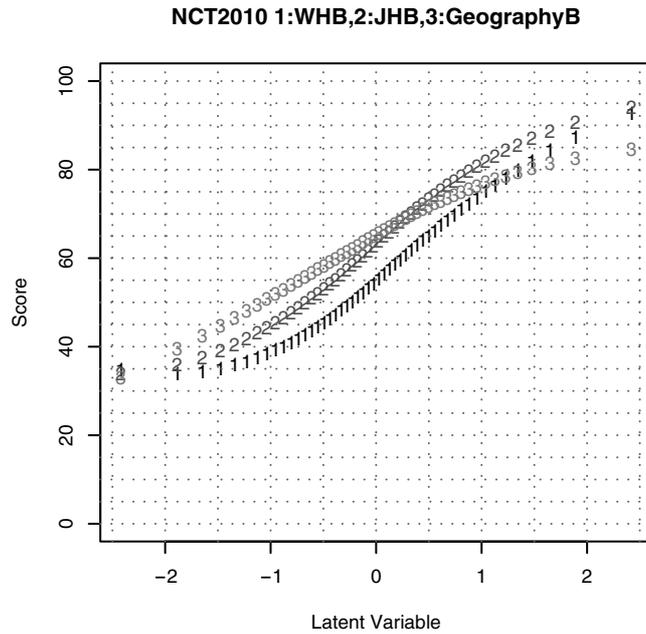


図 10 平成 22 年 1:「世界史 B」, 2:「日本史 B」, 3:「地理 B」得点の条件付平均の推定値
(横軸は潜在変数, 縦軸は科目得点)

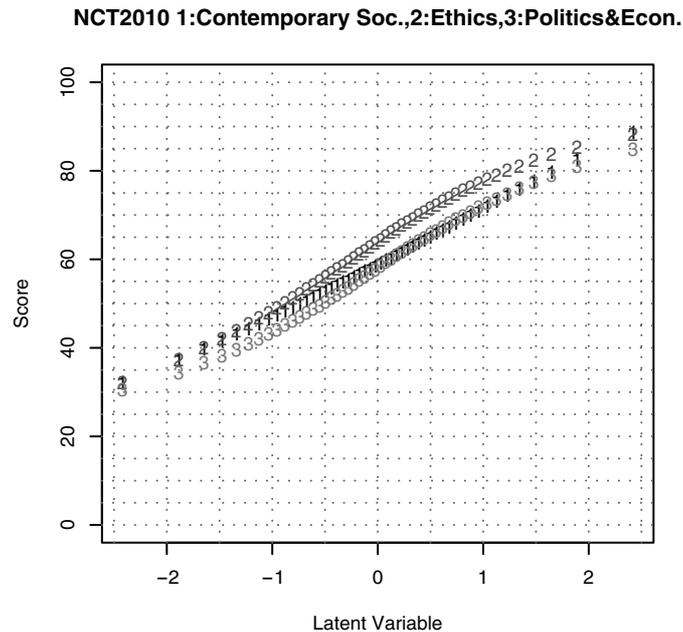


図 11 平成 22 年 1:「現代社会」, 2:「倫理」, 3:「政治・経済」得点の条件付平均の推定値
(横軸は潜在変数, 縦軸は科目得点)

の得点散布図はここでは省略するが、理科科目間の同時受験者についての科目得点の相関係数は、平成 21 年と類似した値を示している (表 12)。このため、特に成績の中位から下の層での、科目得点の推定にはモデルの仮定が厳密には成立しないことの影響があると思われる。

さらに表 11 に示されているように、この分析においては共通性の推定値が科目によってかなり違う。数学と理科の各科目、および英語を対象とした分析では、数学の共通性が大きく推定されており、数学の成績と関係の強い科目得点に大きな重みがかけている。科目のうち、特に「理科総

表 11 分析8 平成 22 年理系科目の特徴

科目	数学 IA	数学 IIB	物理 I	地学 I	理科総合 A	化学 I	理科総合 B
満点	100	100	100	100	100	100	100
本試験受験者	368,289	331,215	147,319	24,406	29,315	208,168	16,372
分析対象件数	364,187	327,209	145,299	23,068	27,270	204,566	14,296
本試験平均	48.96	57.12	54.01	66.76	63.38	53.79	64.83
分析対象平均	49.12	57.33	54.27	67.16	63.38	54.06	65.45
補正済み平均	49.31	55.83	48.42	67.77	66.95	50.34	67.01
共通性の推定値	0.727	0.772	0.698	0.426	0.254	0.663	0.230

科目	生物 I	英語 (+リスニング)
満点	100	250
本試験受験者	184,632	506,898 (リス)
分析対象件数	172,405	364,187
本試験平均	69.70	(118.14+29.39)
分析対象平均	70.47	154.59
補正済み平均	72.06	154.80
共通性の推定値	0.497	0.568

NCT2010 1:PhysI,2:EarthI,3:ChemI,4:Biol

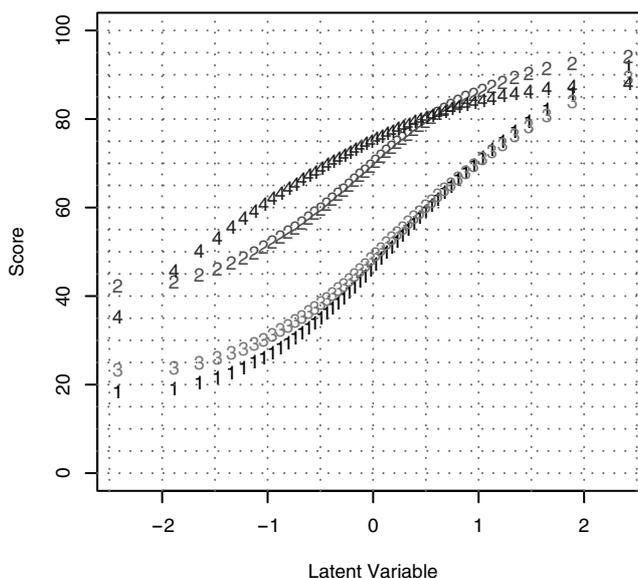


図 12 平成 22 年 1:「物理 I」, 2:「地学 I」, 3:「化学 I」, 4:「生物 I」
得点の条件付平均の推定値

合 A」と「理科総合 B」の共通性の推定値は 0.3 より小さく、他の科目得点との関係が明確ではないことがわかる。また、「地学 I」と「生物 I」の共通性は、「物理 I」・「化学 I」に比べ小さく、いずれも 0.5 に満たない。このような特徴を考慮すると、「地学 I」と「生物 I」の科目得点は、数学成績との関係が薄く、これまでの分析の枠組みでは、特徴を十分にとらえていないために、 $g_j(\eta_j(\theta))$ の形状

が平坦になっていることも考えられる。

5.3 分析 9：平成 22 年数学を除く理系科目の分析

そこで、分析対象とした理系科目のうち「数学 I・数学 A」「数学 II・数学 B」を除いた理科 6 科目と「英語」の科目得点のみを用いて、再度分析を行った (表 13)。共通性の推定値は、「理科総合

表 12 平成 22 年「地学 I」と理科他科目得点との相関係数 (分析対象者)

	物理 I	地学 I	理総 A	化学 I	理総 B	生物 I
物理 I	1.000	NA	0.490	0.731	0.395	0.325
地学 I	0	1.000	0.349	0.234	0.544	0.220
理総 A	3,586	3,589	1.000	NA	0.431	0.373
化学 I	128,026	3,626	0	1.000	0.317	0.533
理総 B	5,980	5,266	2,217	6,154	1.000	NA
生物 I	12,483	6,021	21,369	67,342	0	1.000
分析対象者数	145,299	23,068	27,270	204,566	14,296	172,405

「数学 I・数学 A」, 「英語」および「英語リスニング」の同時受験者のうち対象となる 2 科目を同時受験した者の数と, それらの受験者における 2 科目間の相関を示す. 上三角部分が相関係数, 下三角部分は同時受験者数.

表 13 分析 9 平成 22 年数学を除く理系科目の特徴

科目	物理 I	地学 I	理科総合 A	化学 I	理科総合 B	生物 I	英語 (+リスニング)
満点	100	100	100	100	100	100	250
補正済み平均	52.38	64.34	64.50	52.80	66.83	70.87	154.49
共通性の推定値	0.710	0.595	0.267	0.706	0.399	0.547	0.679

NCT2010 1:PhysI,2:EarthI,3:ChemI,4:Biol

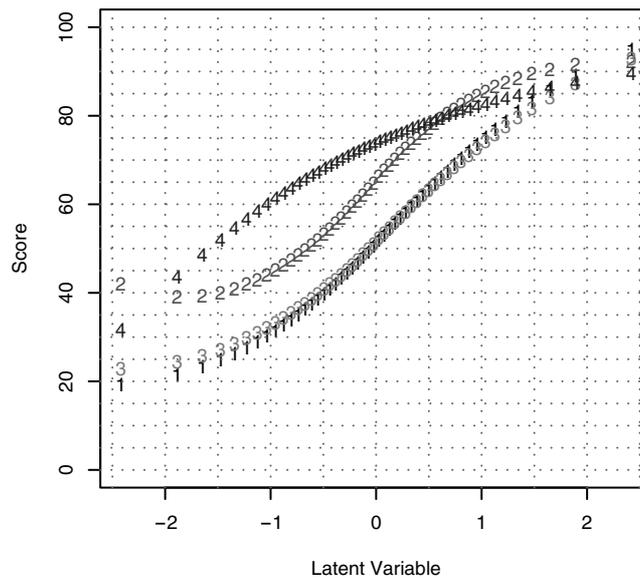


図 13 平成 22 年 1:「物理 I」, 2:「地学 I」, 3:「化学 I」, 4:「生物 I」得点の条件付き平均の推定値 (数学を含まない分析)

A)を除く, 「理科総合 B」, 「地学 I」, 「生物 I」で大きく上昇し, 特に「地学 I」ではほぼ 0.6 に達した. また「化学 I」においても, 共通性が増加し, 「物理 I」においてもわずかではあるが増加が認められる.

得点調整科目間の補正済み平均の違いは, 分析 8 と比較して減少している. 図 13 に示すように, 特に「地学 I」の中位から下位にかけての条件付

き平均が減少しており, 曲線の形状がより急峻になっている. 他方, 「生物 I」の補正済み平均の値は, 分析 8 と比較して減少しているものの, 曲線の形状は大きくは変わっておらず, 学力の中下位層においてなお「物理 I」および「化学 I」と比較して, 大きく条件付き平均の値が異なっていることが分かる. また, 「地学 I」の下位群では, 潜在変数に対する科目得点の条件付き平均が平坦にな

り、一部単調性が崩れている。

平成21年の理系科目の分析(分析7)で示したように、理科の複数科目受験者の同時得点分布は、必ずしも信頼できる情報を与えるとは限らない。このため、数学を分析から除くことにより共通に受験した科目得点についての情報が減少し、歪みを含んだ同時分布の影響が増加する危険も存在するが、数学が他の科目とはかなり異質な性質を持つことに注意を払う必要がある。

6 同時分布に基づく得点調整へ向けて

本稿では、平成20年から平成22年までの3年間について、得点調整科目となっている教科科目得点の同時分布について検討を加えた。すくなくともこの3年間については、理科の科目得点において年度間の変動が大きい。しかしながら、より過去に遡ると、歴史・地理や公民の科目において、調整対象科目間での平均値差が大きくなったこともあり、分布の変動は理科科目に限った現象ではない。すでに大学入試センターによる全国的な大学入学試験は、30回を超えており、その問題作成にあたっては多くの専門家のレビューが毎回行われている。それでもなお時折、得点調整対象科目の平均得点に大きな差が生じるのは、事前テストを用いない難易度調整の限界を示しているように思われる。

現状での大学入試センターが予告している得点調整は、各得点調整科目の周辺分布に基づくものであり、平均点差が20点以上であり、かつその得点差が試験問題の難しさの違いに基づく判断される場合には、平均点差が最大15点になるように調整を行うとしている。実際、比較的最近において科目間の得点差が20点に近い場合を検討すると、受験者層の違いによる科目間の得点の違いと、問題の難しさによる科目間の得点の違いとが同じ方向へ作用している。このため、調整後の平均得点差を0とはせず、平均点差を幾分残すという「保守的」な調整方法には妥当性がある。しかしながら、少なくともここに示した分析から推定すると、分析対象者の科目得点平均と補正済み平均との点差は、受験者が多数存在する科目については、多くとも100点中6点ほどであり、調整対象科目の平均値間で生じた補正の最大が10点を超

えることはない。補正済み平均は、もし分析対象者全員が受験したら得られるであろう科目得点の平均を、MARの仮定のもとで推定するものである。これらの補正量は受験者層の違いによって生じた得点差の推定値とみなせる。このため、現状の制度のもとでの科目得点を考える限り、得点調整対象科目間において平均点で15点の違いが受験者の学力層の違いによって生じるとは考えづらい。

また、さらに重要な問題をはらんでいるのは、受験者層の違いによる科目間の得点差と、試験問題の難しさによる科目間の得点差とが相殺する方向に作用している場合である。このようなケースについては、個別得点の周辺分布に基づく調整方法をとる限り問題を検知することができず、かなり重大な難度の違いがあるにも拘わらず、これを見過ごす可能性が存在する。

昭和54年(1979年)から平成元年(1989年)まで実施された、共通第1次学力試験においては、受験者が受験すべき科目数が一定であったため、英語や数学をアンカーに用いた科目得点の調整を行うことに問題が少なかった。しかし、センター試験においては、受験者の科目選択がきわめて多様であるため、アンカー科目を用いる得点調整が運用上のリスクを考えると実施が困難であるとみなされ、周辺分布のみに基づく得点調整の方法が採用されたと思われる(真弓他, 1999)。しかしながら、最近の情報処理システムの性能向上を考慮すると、分析5に示したような問題はあるものの、科目得点のより多角的な情報を十分に利用した得点調整の方法を考案し実施することは不可能ではない。

ここで利用した分析モデルは、一般的なデータ解析の用途を意図しているために、得点調整の目的には、いささか複雑すぎる性質を持っている。現在のところ、本稿に示したモデル推定のためには、C言語で記述された尤度関数を組み込んだR-2.8システム(R development core team, 2008)を用いて計算を行っている。PCサーバをもちいたシングルスレッドの計算で、現在のところ数時間を一回の推定に要しているが、実装の工夫によりかなりの計算速度の向上は期待し得る。また、得点調整に限定した得点の変換を目的とするならば、石塚・前川(2009)で提案されているように、0点

と満点を固定点とし、変換の関数族はシンプルな区分線形（折れ線）を採用するなど、より単純な方式をとったほうが実用上は望ましい。また、分析5で示された準備の十分でない受験者の存在については、あらかじめ特定の点数基準を設定しておき、各科目について、その基準以上の得点を得た者のみに基づいて変換関数を推定するなどして、これらの受験者の影響を除くなどの工夫をする必要がある。

注

- 1) 一般知能は英語文献では *g*-factor と記述されることが多い。知能測定についての（一般知能論よりの立場からの）簡潔な解説は Deary (2001) にある。また、幾分古い伊藤他 (1981) は知能論全般についての日本の研究者による解説である。
- 2) 大津・林 (2007) および大津 (2009) では独自因子の変動を無視し、潜在変数の分布のみによって科目得点平均の補正値を求めている。
- 3) Available case analysis（観測されている値の対のみを用いてそれぞれの変数の対について相関を求める方法）など
- 4) 潜在変数 θ は標準正規分布に従うことを仮定しているが、この仮定のもとでは $\theta = 1$ は 84.1%分位点にあたり、 $\theta = 2$ は 97.7%分位点にあたる。

謝 辞

2名の査読者のご指摘により、本稿の内容が大きく改善したことに感謝します。

本稿に示された見解は著者個人のものであって、独立行政法人大学入試センターの公的な見解を反映するものではない。また、本稿の内容は、大学入試センター研究開発部内での、得点調整に関する議論と資料分析（石塚・大津・石岡・荘島・橋本・大久保, 2009a, 2009b）に多くを依拠しているが、主張の責任は著者にある。

本研究の実施にあたっては、大学入試センター研究開発部プロジェクト研究 III 「試験問題関連情報の電子的組織化と高機能検索法の研究」、科学研究費補助金基盤研究 (C) (10203829) 「心理学研究における多階層の特性構造を持つデータのための統計モデルの開発」（研究代表者 大津起夫）、および JST/CREST 「現代の産業社会とグレブナー基底の調和」（研究代表者 日比孝之）による支援を部分的に受けた。

参考文献

- 石塚智一・大津起夫・石岡恒憲・荘島宏二郎・橋本貴充・大久保智哉 (2009a). 平成 24 年度以降大学入試センター試験科目の選択方法の変更に伴う得点調整方法の変更試案 (1). 大学入試センター研究開発部リサーチノート RN-09-13.
- 石塚智一・大津起夫・石岡恒憲・荘島宏二郎・橋本貴充・大久保智哉 (2009b). 平成 24 年度以降大学入試センター試験科目の選択方法の変更に伴う得点調整方法の変更試案 (2). 大学入試センター研究開発部リサーチノート RN-09-14.
- 石塚智一・前川眞一 (2009). 区分線形関数による得点調整. 大学入試センター研究紀要 38, 1-11.
- 伊藤隆二・苧阪良二・東 洋・岡本夏木・板倉聖宣・麻生 誠 (1981). 講座 現代の心理学 4 知能と創造性. 岩波書店.
- 大津起夫 (2004). 潜在変数の区分多項式変換を用いた非線形因子分析. 行動計量学 21, 1-15.
- 大津起夫 (2009). 平成 20 および 21 年度大学入試センター試験における科目別得点の難度比較. 大学入試センター研究開発部 リサーチノート RN-09-11.
- 大津起夫・橋本貴充・荘島宏二郎・石塚智一 (2006). 平成 16 年度大学入試センター試験前期日程における志願状況と科目別得点と特徴について. 大学入試センター研究紀要 35, 109-129.
- 大津起夫・林篤裕 (2007). 平成 19 年度大学入試センター試験における科目別得点の難度比較. 大学入試センター研究開発部 リサーチノート RN-06-18.
- 荘島宏二郎・石塚智一・橋本貴充・大津起夫・田栗正章 (2006). 多重指標モニタリングによる得点調整手続きの提案. 大学入試センター研究開発部リサーチノート RN-06-04.
- 橋本貴充・大津起夫・石岡恒憲 (2009a). 平成 20 年度大学入試センター試験の成績ファイル・志願者マスタ・提供者ファイルの CD-R 化作業と主要コードのまとめ. 大学入試センター研究開発部 リサーチノート RN-08-19.
- 橋本貴充・大津起夫・石岡恒憲 (2009b). 平成 21 年度大学入試センター試験データの CD-R 化作業と主要コードのまとめ. 大学入試センター研究開発部 リサーチノート RN-09-01.
- 橋本貴充・大津起夫・石岡恒憲 (2011). 平成 22 年度大学入試センター試験データの CD-R 化作業と主要コードのまとめ. 大学入試センター研究開発部 リサーチノート RN-10-07.
- 真弓忠範・村上隆・白旗慎吾・吉村功・前川眞一 (1999).

大学入試センター試験の得点調整. 大学入試フォーラム 21, 4-18.

de Boor, C. (2001). *A practical guide to splines, Rev. ed.*, New York: Springer-Verlag.

Deary, I.J. (2001). *Intelligence: A very short introduction*. Oxford University Press (ディアリ, イアン 繁樹算男 (訳) 松原達哉 (解説) (2004) 一冊でわかる 知能, 岩波書店)

Gelman, A., Carlin, J.B., Stern, H.S., & Rubin, D.B. (2004). *Bayesian data analysis (2nd ed.)*. Boca Raton, Florida: Chapman & Hall/CRC.

R development core team (2010). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rubin, D.B. (1976). Inference and missing data. *Biometrika*, **63**, 581-592.

A モデルの尤度について

モデルの推定に用いた, 欠測を伴う非線形因子モデルの尤度関数について以下で説明する.

まず, 線形の因子分析と同様に, 潜在変数 θ が標準多変量正規分布 $N(0, I)$ に基づいて分布するものと仮定する. また, 数値的な取り扱いの容易さと表現の多様さの両方を満たすために, η_j としてスプライン関数 (接続点における滑らさの制約を持つ区分多項式) を用いる. さらに, モデル同定の難しさに対応するため, パラメータに緩やかな情報を持つ事前分布を仮定する. ただし, 本文中の分析例ではデータの件数が多いため, パラメータの事前分布を設定することによる効果は小さい. 数値計算においては, 潜在変数についての積分を求めるために, 正確な数値積分を行わず θ の分布を有限個の値をとる離散的な分布によって近似する.

ここで, 観測されたデータが $\{\mathbf{y}_i | i = 1, \dots, n\}$ であり, 各 \mathbf{y}_i の要素が (y_{i1}, \dots, y_{ip}) であるとし, 潜在変数 θ を固定した場合, これらは独立に同一の正規分布に従うとする. このときモデルの尤度は次の式で求められる. ここで尺度付き逆カイ 2 乗分布 (Gelman et al., 2004) $\text{Inv}\chi^2(\psi_j | \nu, s_j^2)$ は, 独自因子の分散 ψ_j が過小な値になることを妨げるためのペナルティである.

$$L(\boldsymbol{\beta}, \boldsymbol{\psi} | \mathbf{y})$$

$$= \prod_{j \in C} \text{Inv}\chi^2(\psi_j | \nu, s_j^2) \times \prod_{i=1}^n \left\{ \sum_{k=1}^K \prod_{j=1}^p p(y_{ij} | \eta_j(\theta_k | \boldsymbol{\beta}_j), \psi_j) w(\theta_k) \right\}. \quad (5)$$

式中の $\boldsymbol{\beta}_j$ は η_j の形状を定めるパラメータの集合を表し, $\boldsymbol{\beta}$ によってそれらのパラメータの全体を表すことにする. 各 η_j はスプライン関数 (区分多項式) であり, 節点 (多項式の区間の端点) はあらかじめ定まっており, 推定されるべきパラメータには含まれないものとする. また, C は変数の集合を表す. 事前分布の自由度 ν を変数に応じて個別に設定することも可能であるが, 本文中では同一の値を用いている (すべての変数について $\nu = 2$ とし, s_j^2 は標本分散に設定した).

確率変数 $\{Y_j\}$ のうち, 観測されている値の集合を \mathbf{y}_o とし, 欠測している値の集合を \mathbf{y}_m で表す. ここで, 観測変数と同様に欠測の指標 $\{I_j\}$ を確率変数と考え, これらの全体, すなわち (I_1, \dots, I_p) を I で表すものとし, その確率分布を $p(I)$ とする. 一般的に I は \mathbf{y}_o と \mathbf{y}_m の両者に依存する可能性があるが, MAR の仮定のもとでは I が \mathbf{y}_m には依存せず, \mathbf{y}_o のみによって分布が定まるものとする. つまり,

$$p(I | \mathbf{y}_o, \mathbf{y}_m, \phi) = p(I | \mathbf{y}_o, \phi)$$

が成立しているとする. ここで ϕ は欠測発生の分布にかかわるパラメータである. 各 Y の分布を定めるパラメータを ζ とする. 利用可能なのは \mathbf{y}_o と I の値であり, これらの確率分布は

$$p(\mathbf{y}_o, I | \zeta, \phi) = \int p(\mathbf{y}_o, \mathbf{y}_m, I | \zeta, \phi) d\mathbf{y}_m = \int p(\mathbf{y}_o, \mathbf{y}_m | \zeta) p(I | \mathbf{y}_o, \mathbf{y}_m, \phi) d\mathbf{y}_m$$

で定義される. ここで MAR の仮定が成立しているならば,

$$p(I | \mathbf{y}_o, \mathbf{y}_m, \phi) = p(I | \mathbf{y}_o, \phi)$$

であり, 上の式は

$$p(\mathbf{y}_o, I | \zeta, \phi) = \int p(\mathbf{y}_o, \mathbf{y}_m | \zeta) d\mathbf{y}_m \times p(I | \mathbf{y}_o, \phi)$$

となる. ここで求めたいのは ζ の値であるが, ζ についての尤度は

$$\int p(\mathbf{y}_o, \mathbf{y}_m | \zeta) d\mathbf{y}_m$$

によって定まり、欠測メカニズム $p(I|\mathbf{y}_o, \mathbf{y}_m, \phi)$ を考慮する必要がない。

MAR の仮定を因子分析の尤度に当てはめると、固定した潜在変数について Y_j が互いに独立なので、次のような簡単な式になる。

$$\begin{aligned} L(\beta, \psi | \mathbf{y}_o) &= \int \prod_{j \in o} p(y_j | \eta_j(\theta | \beta_j), \psi_j) \\ &\quad \times \int \prod_{j \in m} p(y_j | \eta_j(\theta | \beta_j), \psi_j) d\mathbf{y}_m f(\theta) d\theta \\ &= \int \prod_{j \in o} p(y_j | \eta_j(\theta | \beta_j), \psi_j) f(\theta) d\theta. \end{aligned} \quad (6)$$

これは、観測されている変数についてのみ積分を行い、欠測している変数については計算を行う必要がないことを意味する。さらにこれを (5) に適用すると

$$\begin{aligned} L(\beta, \psi | \mathbf{y}) &= \prod_{j \in C} \text{Inv}\chi^2(\psi_j | \nu, s_j^2) \\ &\quad \times \prod_{i=1}^n \left\{ \sum_{k=1}^K \prod_{j \in o_i} p(y_{ij} | \eta_j(\theta_k | \beta_j), \psi_j) w(\theta_k) \right\} \end{aligned} \quad (7)$$

となる。この尤度を最大化する β と ψ を求めることによりモデルの推定を行う。

B スプライン関数について

k 次のスプライン関数は次のような区分的多項式の和として定義される。ここで $t_0 < t_1 < \dots < t_{n+1}$ を節点（多項式を接続する点）の集合とする^{注5}。関数は節点で定義される各区間 $[t_j, t_{j+1}]$, ($j = 0, \dots, n$) において k 次多項式であり、また節点において $k-1$ 次までの微分が連続であるものに限定する。記号 $(x)_+$ によって x が正のとき値 x をとり、ゼロ以下のときにはゼロである関数を表す。 $P_k(x)$ を任意の k 次多項式とすると

$$S(x) = P_k(x) + \sum_{j=1}^n a_j (x - t_j)_+^k \quad (8)$$

によって区分的な k 次多項式が定義できる。これが k 次スプライン関数と呼ばれるものである。式 (8) の和記号中の各項 $(x - t_j)_+^k$ は、それぞれ $x = t_j$ において $k-1$ 次までの微分が連続であるので、上に述べた条件を満たしていることがわかる。これにより P_k の $k+1$ 個の係数と $\{a_j | j = 1, \dots, n\}$ を用いて多様な関数を表現することができる。特に $k=3$ とおいた場合はキュービックスプラインと呼ばれ、利用される頻度が多い。また $k=1$ とすると折れ線になる。

スプラインの定義 (8) を少し変更することにより、滑らかさの制約を弱めた関数を表現することも可能である。具体的には制約を弱める節点については、 $(x - t_j)_+^{k'}$, ($k' < k$) の項を加える。このとき節点 t_j では $k'-1$ 次までの微分のみが連続となり、滑らかさの制約が緩められる。これらは拡張スプラインと呼ばれる。微分の連続性を弱める場合には、形式的にスプラインを定義する節点が重複しているものとして表記する。

式 (8) は切断べき関数による表現とよばれるが、これを用いて関数の形を定めようとする、項が互いに一次従属に近い場合しばしば数値的な悪条件に陥る。

切断べき関数と理論的に同等な関数を定義し、より数値的に安定した表現を与えるものとして広く利用されているのが B スプラインである。 B スプラインは、(1) 関数を表現するための基底を区分多項式によって表現する、(2) 各基底が非ゼロの値を持つのは局所的な領域に限られる、という特徴を持つ。

本文中の分析では、潜在変数 θ の関数 $\eta_j(\theta)$ を内部節点が1個のスプライン関数とした。これは、実質的には $\theta = 0$ において1次微分の等しい連続に接続する2つの2次多項式によって η_j を表すことになる。この場合、関数 η_j を定めるためには4つのパラメータが必要になる。

注

5) de Boor の用法では k 階 (order k) のスプライン関数とは $k-1$ 次区分多項式を意味する。

Comparison between test scores of NCT subjects with nonlinear factor analysis

OTSU Tatsuo*

Abstract

The author compared the test difficulties between some subjects of NCT in 2008, 2009 and 2010. The NCT had announced that scores of the subjects would be adjusted in case of large mean deviances from the other subjects in the same area.

A method of nonlinear factor analysis (NFA) for incomplete data (Otsu,2004) was used for analyzing the test scores. Classical factor analysis model represents linear relationships between latent variables and observed variables. Although this is widely used for analysis of psychological tests, nonlinear relationships need to be represented for analyzing scores of large scale achievement tests.

Here, a nonlinear factor analysis model that uses spline transformation of latent variables is applied for analyzing the NCT scores. The conditional distribution of observed variables given by the latent variables is assumed to have means (or location parameters) that are expressed in nonlinear transformations of the latent variables. The discrete approximation of the latent variables enables easy adaptation for the missing values under MAR (missing at random) condition.

It is suggested that the mean score differences between marginal distributions are insufficient for monitoring the test difficulties. The whole shape of the simultaneous score distribution should be considered for good adjustment.

Key words: NCT, test difficulty, score adjustment, nonlinear factor analysis, MAR

* The National Center for University Entrance Examinations
JST/CREST