

本追モニター調査の結果を利用した 傾向スコア加重法による本追試験間の難易度比較

吉村 宰*
 莊島宏二郎*

要 約†

本稿では、大学入試センター試験及び法科大学院適性試験のモニター調査参加者のデータから、本試験受験者が追試験を受験した場合の平均点や得点分布を推定する方法及びその実データへの適用例を報告する。

キーワード：傾向スコア，直接修正，加重法，欠損値，大学入試センター試験，法科大学院適性試験，モニター調査

1 はじめに

大学入試センター試験は毎年、本試験と追試験の2回の試験が行われている。両試験とも区別なく同一年度の大学入学者選抜に利用されることを考えると、その難易度は理想的には同一であることが望ましい。また、平成15年度からは法科大学院適性試験も実施されるようになったが、これも同一年度内に本試験と追試験の2回の試験実施が予定されており、両試験の難易度に大きな違いが生じないことが期待されている。

試験で用いられる各項目の難易度を事前に調査できるのであれば、2つの異なる試験をその難易度を等しく構成することにそれほどの困難はない。しかし、我が国の大学入学者選抜試験を巡る社会・文化的背景は、試験問題の事前調査を許容するものでなく、本試験と追試験との難易度が大きく異なるよう試験問題作成者による「多大な努力」が払われているのが現状である。このような状況で、受験者と試験問題の両方が異なる本試験と追

試験の難易度を比較することは、原理的に不可能である。

大学入試センターでは、毎年、大学1年生400名を対象とした、センター試験のモニター調査を行っている。これは、同一集団に本試験と追試験の両方を受験させることで難易度の比較検証を行い、次年度以降の作題の参考資料を提供しようとするものである。平成15年度モニター調査の対象となった科目は、本試験受験者数の多い、以下の6教科14科目である（林，2003）。

- 国語：国語Ⅰ・Ⅱ
- 地歴：世界史B，日本史B，地理B
- 公民：現代社会，倫理，政治経済
- 数学：数学Ⅰ・数学A，数学Ⅱ・数学B
- 理科：物理IB，生物IB，化学IB，地学IB
- 外国語：英語

法科大学院適性試験についても、都内4大学に在学する3年生計約400名を対象にモニター調査が行われている。目的は大学入試センター試験におけるモニター調査と同じである。

表1，表2にそれぞれ、平成15年度大学入試セ

* 大学入試センター研究開発部 試験作成支援研究部門
 〒153-8501 東京都目黒区駒場 2-19-23 osamu@rd.dnc.ac.jp
 2003年11月20日 受理

表 1 実受験者とモニターの平均得点及び標準偏差 (平成 15 年度大学入試センター試験国語 I・II)

	受験者数	平均値	標準偏差
実受験者 (本試験)	471,980	101.08	28.84
実受験者 (追試験)	268	111.64	31.08
モニター (本試験)	350	120.42	28.88
モニター (追試験)	350	122.40	30.36

表 2 実受験者とモニターの平均得点及び標準偏差 (平成 15 年度法科大学院適性試験)

	受験者数	平均値	標準偏差
実受験者 (本試験)	28,325	63.07	14.19
実受験者 (追試験)	7,174	59.80	14.77
モニター (本試験)	321	65.59	14.78
モニター (追試験)	321	64.23	15.44

ンター試験国語 I・II 及び平成 15 年度法科大学院適性検査の本試験及び追試験における実受験者とモニター調査参加者の平均得点及び標準偏差を示した。

表 1 から分かるように、実受験者では追試験の方が本試験よりも約 10 点平均点が高いが、モニター調査参加者ではその差は 2 点程度である。また、法科大学院適性試験では、実受験者で追試験の方が本試験よりも約 3 点低いが、モニター調査参加者ではその差は 1 点程度しかない (表 2)。

こうしたことを根拠に、平成 15 年度大学入試センター試験国語 I・II 及び法科大学院適性試験の本試験と追試験の間には顕著な難易度の差はないと主張することができる。しかし一方で、モニター調査参加者の平均点は本試験で約 20 点、追試験で約 10 点実受験者よりも高い。センター試験モニター調査参加者は実受験者に比べ高い学力の集団であることが明らかである。学力中位層はよく識別するが、上位層の学力の識別性はそれほど良くないというセンター試験の特性を考慮すると、仮に本試験の実受験者が追試験を受けた場合の本試験と追試験の平均点の差が、モニター調査参加者に見られた平均点の差よりも大きくなる可能性もあり、追試験は本試験よりも易しかったのではないかという批判あるいは疑念も起こり得る。このことは、法科大学院適性試験についても同様である。

モニター調査の対象者が実受験者の学力分布を代表するように構成されているならば、上記のよ

表 3 データの構造

	Y	Z	$\mathbf{x}$		
			x_1	$\cdots$	x_m
1	y_{11}	1	x_{11}	$\cdots$	x_{m1}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	y_{1i}	1	x_{1i}	$\cdots$	x_{mi}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	y_{0k}	0	x_{1k}	$\cdots$	x_{mk}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	y_{0N}	0	x_{1N}	$\cdots$	x_{mN}

うな批判や疑念は起こらない。最善は本試験受験者が追試験も受験することである。次善は、本試験受験者を代表するような無作為標本を構成し、これを対象としたモニター調査を行うことである。しかし、これらはいずれもコスト等の面から現実的ではない。

学力分布が実受験者集団とは明らかに異なる集団を対象としたモニター調査の結果を利用しつつ、かつ「仮に」本試験の実受験者集団 (と同質の集団) が追試験を受験した場合の追試験の平均値や標準偏差及び得点分布等がどうなるかを知ることができる方法が必要である。

本稿では、傾向スコア加重修正法によってモニター調査参加者の本試験の得点分布を実受験者集団のそれに近似し、加重修正されたモニター調査参加者の追試験の得点分布と実受験者による本試験のそれとを比較することにより、本試験と追試験の難易度の比較を試みる。

2 傾向スコア加重法による欠損群の平均値の推定

個体 i ($i = 1, 2, \dots, N$) から変数 Y が観測されることを $z_i = 1$ 、観測されないことを $z_i = 0$ とし、個体 i から変数 Y が観測された場合を Y_{1i} 、観測されない場合を Y_{0i} と表現する。また、 Y が観測された個体の数を n_1 、観測されない個体の数を n_0 とする ($n_1 + n_0 = N$)。なお、すべての個体からは、 Y に関するすべての共変量 $\mathbf{x}$ が観測されているものとする (表 3)。

このとき、ignorable selection の仮定のもとで、全個体から Y が観測されたときの Y の期待値 $E(Y)$ は以下で求められる (Rosenbaum, 1987)。

$$E(Y) = \frac{1}{N} \sum_{i=1}^N \frac{z_i Y_{1i}}{\hat{e}(\mathbf{x}_i)} \quad (1)$$

ただし $\hat{e}(\mathbf{x}_i)$ は $e(\mathbf{x}_i) = Pr(z_i = 1|\mathbf{x}_i)$
(傾向スコアの推定値)

これを傾向スコア加重修正法 (the propensity score weighting) という。傾向スコアについての詳細は文献 Rosenbaum & Rubin (1983a, 1983b, 1984), Rosenbaum (1984, 2002), Imbens (2000) など参照されたい。なお、傾向スコア $e(\mathbf{x})$ の推定値は、 $Z(=1,0)$ を目的変数、共変量 $\mathbf{x}$ を説明変数としたロジスティック回帰で推定された予測値を用いて得られることが多い。すなわち、

$$\hat{e}(\mathbf{x}_i) = \frac{1}{1 + \exp\{-(\hat{\beta}_0 + \hat{\beta}\mathbf{x}_i)\}} \quad (2)$$

ところで、式 (1) の $E(Y)$ は、欠損群と観測群を合わせた集団における Y の平均である。これは本稿の文脈で言うと、本試験の実受験者とモニターとを合せた集団の平均となる。本稿の目的は、本試験の実受験者集団が追試験を受験した場合の得点の平均値を得ることにある。そのためには $E(Y_0)$ を求める必要がある。

Hahn (1998) は Y を Y_1, Y_0 を用いて、

$$Y_i \equiv z_i Y_{1i} + (1 - z_i) Y_{0i} \quad (3)$$

と表現している。これを用いると Y の期待値は

$$E(Y) \equiv \frac{1}{N} \sum_{i=1}^N \{z_i Y_{1i} + (1 - z_i) Y_{0i}\} \quad (4)$$

となる。

ここで、 Y_0 の期待値 $E(Y_0)$ を考える。これは、欠損群における n_0 個の Y_0 の標本平均として得られる。すなわち、

$$E(Y_0) = \frac{1}{n_0} \sum_{i=1}^{n_0} Y_{0i} \quad (5)$$

Y_0 は観測されていないが、式 (1)(3) を利用すると、

$$\begin{aligned} E(Y_0) &= \frac{1}{n_0} \sum_{i=1}^{n_0} Y_{0i} \\ &= \frac{1}{n_0} \left\{ \sum_{i=1}^N \frac{z_i Y_{1i}}{\hat{e}(\mathbf{x}_i)} - \sum_{i=1}^N z_i Y_{1i} \right\} \end{aligned}$$

表 4 センター試験モニター調査におけるデータの構造

		追試験	Z	本試験		
				項目 1	...	項目 m
モニター	1	80	1	1	...	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
モニター	350	65	1	0	...	1
実受験者	351	missing	0	1	...	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
実受験者	3350	missing	0	1	...	0

$$\begin{aligned} &= \frac{1}{n_0} \sum_{i=1}^N \left\{ \frac{z_i Y_{1i}}{\hat{e}(\mathbf{x}_i)} - z_i Y_{1i} \right\} \\ &= \frac{1}{n_0} \sum_{i=1}^N \frac{z_i Y_{1i} (1 - \hat{e}(\mathbf{x}_i))}{\hat{e}(\mathbf{x}_i)} = \bar{Y}_0 \quad (6) \end{aligned}$$

が得られる。また、欠損群における Y の分散 $V(Y_0)$ も同様に、

$$\begin{aligned} V(Y_0) &= \frac{1}{n_0 - 1} \sum_{i=1}^{n_0} (Y_{0i} - \bar{Y}_0)^2 \\ &= \frac{1}{n_0 - 1} \left\{ \sum_{i=1}^N \frac{z_i Y_{1i}^2 (1 - \hat{e}(\mathbf{x}_i))}{\hat{e}(\mathbf{x}_i)} - \frac{n_0}{n_0 - 1} (\bar{Y}_0)^2 \right\} \quad (7) \end{aligned}$$

として得られる。

3 平成 15 年度大学入試センター試験モニター調査データへの適用

3.1 データ

平成 15 年度センター試験国語 I・II 全受験者から無作為に抽出した 3000 名分の成績データ、及び平成 15 年度センター試験モニター調査協力者 350 名の本試験及び追試験の成績データを用いた。これを表 3 と同様に表現すると表 4 のようになる。

センター試験本試験の実受験者の追試験の成績は観測されていないが、本試験の成績データを共変量 $\mathbf{x}$ とし、傾向スコア $e(\mathbf{x}_i) = P(Z_i = 1|\mathbf{x}_i)$ を用いモニター調査参加者の追試験データを加重修正することによって (式 (5)(6) の利用)、モニター調査参加者の本試験得点分布が実受験者のそれとほぼ同じであった場合の、追試験の平均、標準偏

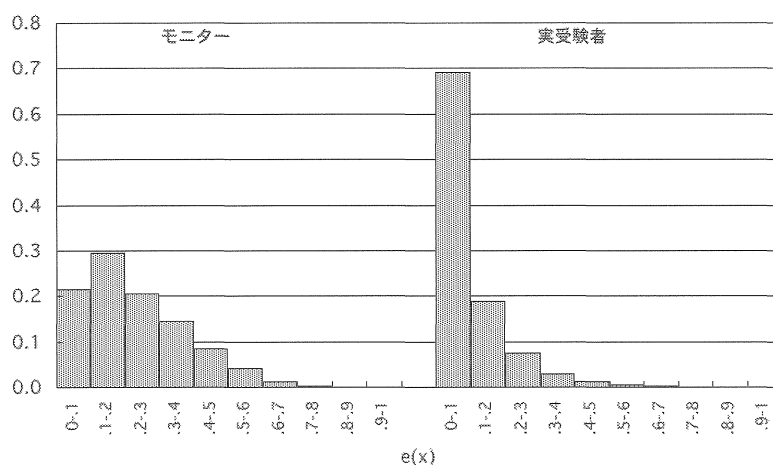


図 1 傾向スコアの推定値の分布

差及び得点分布を求めることができる。

3.2 傾向スコアの推定

追試験受験の有無を表すダミー変数 $Z(= 1, 0)$ を目的変数、本試験の全項目への正誤データ及び得点の五段階評価データを説明変数 x としたロジスティック回帰分析によって、傾向スコアの推定値 $\hat{e}(x_i)$ を求めた。モニター群と実受験者群それぞれの傾向スコアの推定値の分布を図 1 に示す。なお、傾向スコアの推定に用いた共変量は次の通りである。

- 項目 1～4, 6, 7, 12, 13, 15～17, 21, 23～28, 30～35 への正誤データ
- 得点を 40 点間隔で 5 分割したとき、どの範囲に属するかを表す二値データ

また、傾向スコアの推定に用いる共変量は次の手順で決定した。

1. 共変量の候補（項目への正誤反応や総得点）を選ぶ。
2. 傾向スコアを推定する。
3. 推定された傾向スコアによる加重修正後のモニターにおける平均得点と標準偏差、及び項目正答率を算出し、実受験者におけるそれらとの差異を確認する。
4. 実受験者と加重修正後のモニターとの間の平均得点と標準偏差、及び項目正答率の差異が満足できる小ささになるまで 1) から 4) を繰り返す。

この手順は試行錯誤的に行った。

採用した共変量の場合、全項目への正誤反応を共変量とした場合、総得点のみを共変量とした場合、最終的に採用した項目と総得点を共変量とした場合の修正後のモニターにおける平均得点、標準偏差、及び項目正答率(%)の差(絶対値)の平均はそれぞれ次の通りであった。なお、本試実受験者における平均得点は 100.30、標準偏差は 28.96 であり、加重修正前のモニターとの項目正答率の差(絶対値)の平均は 11.4%であった。

採用した共変量 平均得点 = 100.61, 標準偏差 = 29.82, 項目正答率の差の平均 = 2.7%
 全項目への正誤反応 平均得点 = 102.36, 標準偏差 = 19.62, 項目正答率の差の平均 = 3.7%
 総得点のみ 平均得点 = 100.58, 標準偏差 = 26.03, 項目正答率の差の平均 = 6.4%
 採用した項目と総得点 平均得点 = 99.60, 標準偏差 = 29.80, 項目正答率の差の平均 = 2.6%

図 1 から分かるように、モニター群と実受験者群では傾向スコアの分布は極端に異なる。特に傾向スコアが 0～0.1 である者の割合がモニター群に比べ実受験者群で非常に高い。言い換えると、モニター群では傾向スコアが 0～0.1 の層が少なく、実受験者集団が追試験を受験した場合の得点分布を推定するためには、こうした少ない層に大きな重みを与えるように加重修正する必要がある。これが傾向スコア加重法の本質である。

表 5 モニター加重修正後のセンター試験本試験の平均値と標準偏差の推定値

	平均値	標準偏差
本試験：実受験者（3,000 名）	100.30	28.96
本試験：モニター加重修正による推定値	100.61	29.82

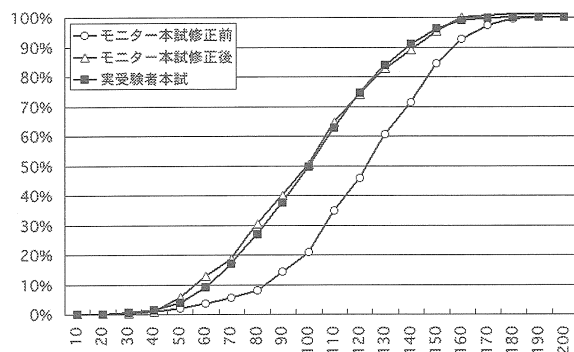


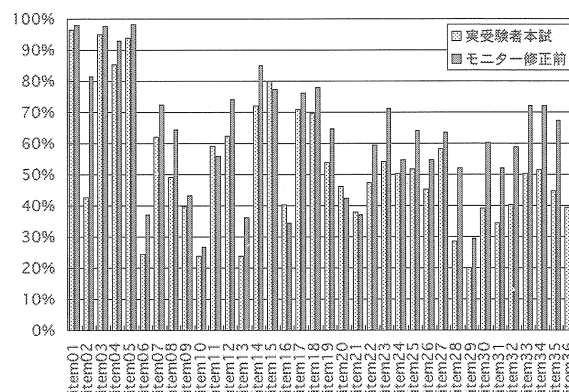
図 2 実受験者及びモニター（加重修正前後）のセンター試験本試験累積得点分布

3.3 モニター調査データの加重修正によるセンター試験本試験の平均値と標準偏差，得点分布

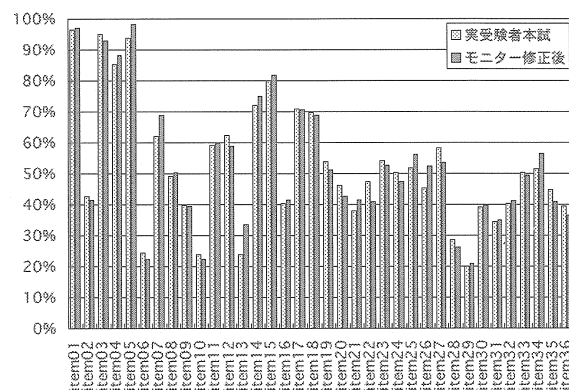
傾向スコアによる加重修正の成否の程度を確認するために，得られた傾向スコアを用いてモニター調査本試験データの加重修正によって実受験者群の本試験得点の平均値，標準偏差，及び得点分布を推定し，実際の実受験者の本試験における得点の平均値，標準偏差，得点分布との比較を行った（表 5，図 2）。

表 5 及び図 2 に示されたように，モニター調査参加者の本試験における得点分布は，傾向スコア加重法によってその偏りが修正され，本試験の実受験者の得点分布にかなり近づいていることが分かる。

図 3 に，加重修正前と加重修正後の実受験者とモニターにおける項目正答率を示した。加重修正前の実受験者とモニターでは多くの項目で正答率に大きな違いが見られるが，加重修正によって，これら項目正答率の違いが大幅に減少していることが分かる。このことから，モニターを加重修正することによって得られる学力分布は，実受験者集団のそれとかなり似たものであると言える。



加重修正前



加重修正後

図 3 加重修正前後の実受験者とモニターの項目正答率の比較

3.4 モニターの加重修正による実受験者近似群におけるセンター試験追試験得点の平均値，標準偏差，得点分布の推定と本追難易度比較

傾向スコア加重法により，モニター調査参加者の本試験の得点分布を実受験者のそれに近似した場合，すなわち，モニター調査参加者の学力分布の偏りを修正した場合の，モニター調査参加者の追試験の平均点，標準偏差，得点分布を求めた。表 6 は，実受験者，モニター及び加重修正されたモニターによる本試験と追試験の平均得点及び標準偏差，図 4 はそれぞれの得点の累積分布である。

表 6 実受験者及びモニター（加重修正前後）におけるセンター試験本追試験の平均点等の比較

	N	平均値	標準偏差
本試験：実受験者	3,000	100.30	28.96
追試験：実受験者	268	111.64	31.08
本試験：モニター	350	120.42	28.84
追試験：モニター	350	122.40	30.36
本試験：モニター加重修正	350	100.61	29.82
追試験：モニター加重修正	350	104.67	30.51

表 6 に示されたように、モニター調査参加者の得点分布を本試験の実受験者のそれに近似した場合の追試験の平均点と実際の本試験の平均点の差（約 4 点）は、モニター調査参加者における本試験と追試験の平均点の差（約 2 点）に比べ若干であるが大きくなっている。モニター調査参加者の学力分布が実受験者に比べ高い方に偏っていることから、実受験者集団における本試験と追試験の平均点の差異は、モニター調査参加者に見られる差異よりも大きくなる可能性があることは先に指摘した通りである。しかし、それがどの程度であるかまでは不明であった。

ここで得られた結果は、仮に本試験の実受験者集団が追試験を受験した場合、その平均点は本試験に比べ約 4 点程度高くなったであろうことを示すものである。おそらく追試験は本試験に比べやや易しいものであり、また、追試験の受験者集団は本試験の受験者集団に比べやや学力が高かったであろうことが分かる。

4 平成 15 年度法科大学院適性試験モニター調査データへの適用

4.1 データ

平成 15 年度法科大学院適性試験本試験受験者 28,325 名分の成績データ、及び同試験モニター調査協力者 321 名の本試験及び追試験の成績データを用いた。

法科大学院適性試験はそれぞれ約 30 項目からなる第 1 部（推論）と第 2 部（読解）から構成されている。試験時間は第 1 部と第 2 部とも 90 分であり、第 1 部と第 2 部の間に 30 分の休憩時間が設定されている。各部の満点はそれぞれ 50 点、第 1 部と第 2 部の合計点の満点は 100 点である。

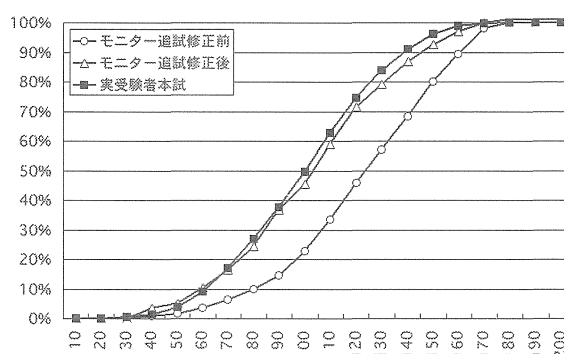


図 4 実受験者及びモニター（加重修正前後）のセンター試験追試験累積得点分布

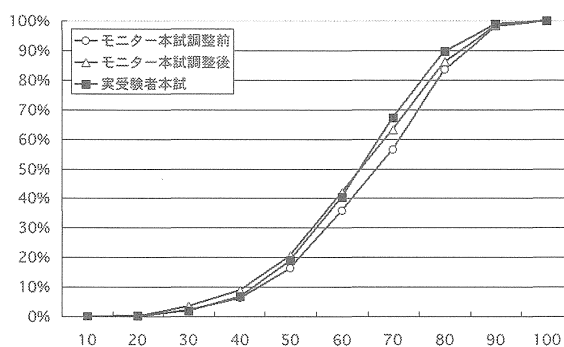


図 5 実受験者及びモニター（加重修正前後）の法科大学院適性試験本試験累積得点分布

4.2 傾向スコアの推定に用いた変数

傾向スコアの推定に用いた変数は次の通りである。

- 第 1 部の得点、第 2 部の得点
- 第 1 部の設問 1,4,5,8,13,14,19 への正誤データ
- 第 2 部の設問 6,9,13,16,22,25 への正誤データ

4.3 モニター調査データの加重修正による法科大学院適性試験本試験の平均値と標準偏差、得点分布

図 5 は法科大学院適性試験本試験における実受験者の総合点と加重修正前後のモニターの総合点の累積得点分布を併せて表示したものである。センター試験ほど顕著ではないが、モニターの得点分布が実受験者に比べ高いことが分かる。傾向スコア加重法によって本試験の得点分布の違いは、かなりの程度修正されている。

図 6、図 7 はそれぞれ第 1 部と第 2 部における加重修正前後の項目正答率である。第 1 部では実

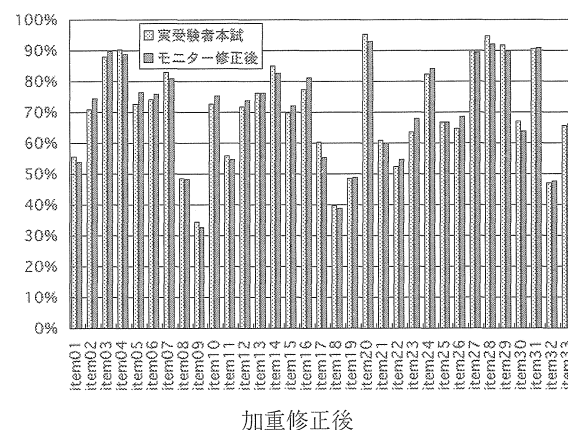
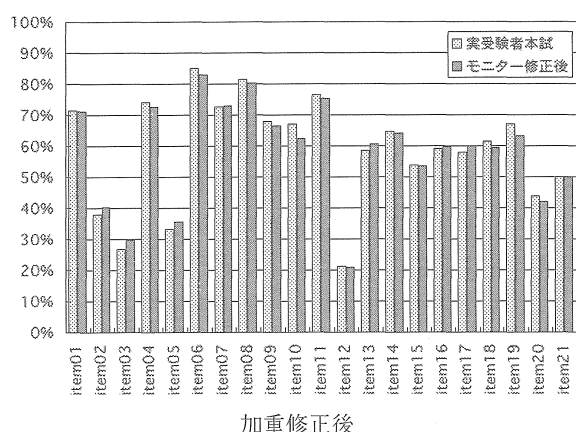
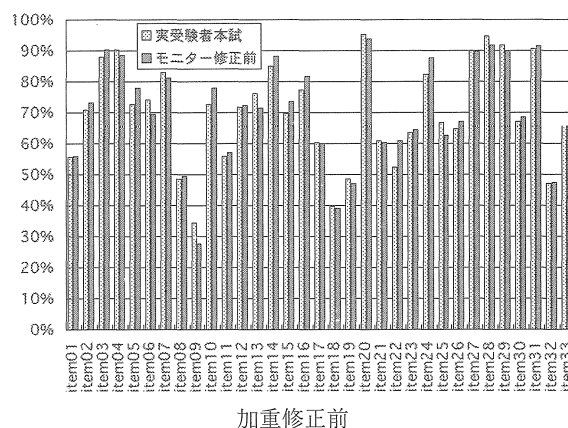
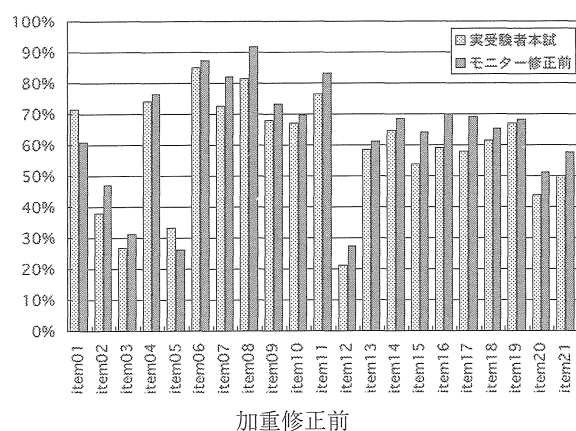


図 6 加重修正前後の実受験者とモニターの項目正答率の比較（第 1 部：推論）

図 7 加重修正前後の実受験者とモニターの項目正答率の比較（第 2 部：読解）

受験者とモニターの項目正答率の違いが比較的顕著であったが、加重修正によりこの違いは大幅に減少していることが分かる（図 6）。第 2 部では加重修正前の項目正答率に実受験生とモニターとの違いがあまり見られなかったが、同じ重みによる修正によって、若干違いは減少している。

図 5～7 からも分かるように、加重修正後のモニターはその得点分布等において実受験者集団をよく近似していると言える。

4.4 モニターの加重修正による実受験者近似群における法科大学院適性試験追試験得点の平均値、標準偏差、得点分布の推定と本追難易度比較

本試験での得点分布等がほぼ同じとなるように求められた重みを使い、モニターにおける追試験データを加重修正することで、本試験受験者集団とほぼ同等の得点分布等をもつ集団が追試験を受験

した場合の追試験の成績を推測することができる。

表 7 に、実受験者、モニター及び加重修正されたモニターによる本試験と追試験の平均得点及び標準偏差を示した。また、図 8 は実受験者、モニター及び加重修正されたモニターによる追試験における累積得点分布である。

実受験者の本試験と追試験を比較すると、追試験の方が約 3 点平均点が低い。世間一般では、こうした平均点の違いの原因を直ちに試験問題の難易度の違いに帰着させる傾向が強い。特に平成 15 年度の法科大学院適性試験では、特例措置によって追試験受験者数が受験生全体の約 20% を占めるまでに至っている。このため試験の難易度の違いに対する関心も高いと思われる。

こうした状況で、特定の大学に所属する学生を対象としたモニター試験の結果を示し、本試験と追試験の間に難易度の違いはないと主張することは、やや説得力を欠く恐れがある。本試験受験者

表 7 実受験者及びモニター（加重修正前後）における法科大学院適性試験本追試験の平均点等の比較

	N	平均値	標準偏差
本試験：実受験者	28,325	63.07	14.19
追試験：実受験者	7,174	59.80	14.77
本試験：モニター	321	65.59	14.78
追試験：モニター	321	64.23	15.44
本試験：モニター加重修正	321	62.90	14.14
追試験：モニター加重修正	321	62.15	15.22

が追試験を受験した場合にはどうなったかという問いに対し具体的な数字を挙げて説明できないからである。

モニター加重修正による本試験と追試験の平均点の比較は、まさにこの問いへの回答となっている。ここで適用した法科大学院適性試験のデータに関して言えば、本試験において実受験者集団と同等の成績の集団が追試験を受験した場合、その平均点の違いは1点未満である、ということを示すことができる。

図8からも分かるように、平成15年度法科大学院適性試験の本試験と追試験の難易度にはほとんど違いがなく、実受験者による追試験の平均点がやや低かったのは、この受験者集団の性質に起因するものであると主張することも可能だろう。

5 おわりに

追試験の受験者数はこれまで少なく、従来、本試験と追試験の難易度の違いに対する関心もそれほど高いものではなかった。しかし、何らかの事故等により大量の追試験受験者が生じる可能性は無いとは言えない。本試験と追試験の難易度比較のためのより説得力のある方法を準備しておくことは必要である。

本試験と追試験の難易度の比較は、本来ならば、実受験者集団において行われることが望ましい。モニターを用いた本試験と追試験の難易度の比較では、モニターの学力分布の偏りが問題となる。本稿で報告した手法を用いると、モニターの学力分布の偏りを修正した上での本試験と追試験の成績の比較を行うことができる。また、本手法には簡便であるという利点がある。あくまでも「実受験者集団と似た学力分布の集団が追試験を受験したならば」という仮定の下での議論であることさ

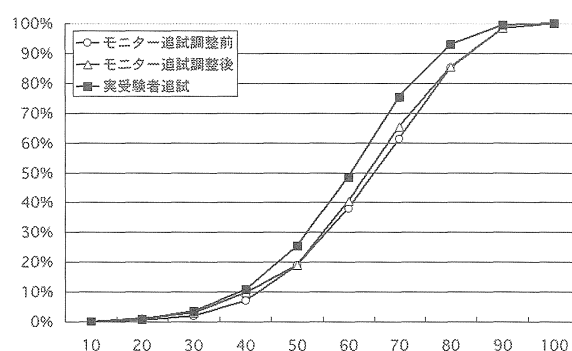


図 8 実受験者及びモニター（加重修正前後）の法科大学院適性試験追試験累積得点分布

え念頭に置いておけば、有用な参考情報となることは間違いのないであろう。

さらに、本稿で報告した手法はIRTの母数推定にも応用できる。例えば、尤度を傾向スコアで重みづけすると、モニター調査の追試験のデータから、本試験の実受験者集団からデータが得られた場合の母数の推定を近似的に行うことができる。

しかし課題も残されている。本稿で示した適用例では、傾向スコア推定のために用いた共変量セットの決定を相当回数の試行錯誤的探索により行った。これは、他のより効率的であると考えられる2つの方法を試みたが、いずれもうまくいかなかったためである。

1つは、傾向スコアを推定する際に用いるロジスティック回帰分析で、2群（実受験者とモニター）とをよりよく判別する共変量を選ぶという方法である。この場合、2群をよく判別する共変量によって推定された傾向スコアが、必ずしも適切な重みとはならないという結果であった。もう1つは、加重修正後のモニターにおける平均得点及び標準偏差が実受験者の平均得点、標準偏差と一致するように、ロジスティック回帰における回帰係数を直接推定するという方法である。この場合、平均得点と標準偏差は一致するが、両集団における項目正答率の差異は改善されないということが分かった。

例えば、上述した後者の方法において、さらに高次のモメントを一致させるように回帰係数を推定するなど、他にも可能性は考えられる。最適な共変量選択の定式化を今後の課題としたい。

参考文献

- Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of Average treatment effects, *Econometrica*, 66, 315–332.
- 林篤裕 (2003). 平成 15 年度大学入試センター試験の本試験と追試験の難易度比較に関する調査研究と標準化の試み, 大学入試センター研究開発部研究成果報告書.
- Imbens, G.W. (2000). The role of the propensity score in estimating does-response functions, *Biometrika*, 87 (3), 706–710.
- Rosenbaum, P.R. (1984). From Association to Causation in Observational Studies: The Role of Test of Strongly Ignorable Treatment Assignment, *Journal of American Statistical Association*, 79, 41–48.
- Rosenbaum, P.R. (1987). Model-Based Direct Adjustment, *Journal of American Statistical Association*, 82, 387–394.
- Rosenbaum, P.R. (2002). *Observational Studies*, second edition, Springer.
- Rosenbaum, P.R. and Rubin, D.B. (1983a). The central role of the propensity score in observational studies for causal effects, *Biometrika*, 70 (1), 41–55.
- Rosenbaum, P.R. and Rubin, D.B. (1983b). Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome, *Journal of Royal Statistical Society, B*, 45 (2), 212–218.
- Rosenbaum, P.R. and Rubin, D.B. (1984). Reducing Bias in Observational Studies Using Subclassification on the Propensity Score, *Journal of American Statistical Association*, 79, 516–524.

Comparing difficulties of a main test and it's supplementary by the propensity score weighting

An application to NCUEE Test and Aptitude Test
for Law School Admissions

YOSHIMURA Osamu*
SHOJIMA Kojiro*

Abstract

In this paper, we demonstrated that the mean and the distribution of the supplementary test scores of the main test takers in each of NCUEE Test and Aptitude Tests for Law School Admissions could be estimated by weighting the score of the examinee of experimental Monitor Supplementary Tests.

Key words: the propensity score, direct adjustment, weighting, missing data, NCUEE Test, Aptitude Test for Law School Admissions

* Department of Applied Statistics and Measurement, Research Division,
National Center for University Entrance Examinations