

離散得点分布の分位点の推定誤差

——スタナイン分位点に着目して——

荘島宏二郎, 石岡恒憲, 大津起夫, 宮澤芳光, 橋本貴充, 前川眞一 (大学入試センター)

令和 3 (2021) 年 1~2 月に実施された第 1 回大学入学共通テストでは, 各試験のスタナインが公表された。スタナインは, 受験者を 9 集団に割合分割した順位数である。スタナインの利用には, スタナイン分位点が精度よく求まっている必要がある。本研究では, テスト得点が離散値であることで, スタナイン分位点が離散化されることの影響と, スタナイン分位点の点推定値と区間推定の推定精度を数値実験によって検討することである。数値実験の結果, 分位点の点推定値はブートストラップ法, 分位点の標準誤差の推定はカーネル密度推定が最も精度が高かった。

キーワード: スタナイン, 分位点, 標準誤差, カーネル密度推定, ブートストラップ

1 はじめに

令和 3 (2021) 年 1~2 月に実施された第 1 回大学入学共通テスト (以下, 共通テスト) より, スタナインが公表されるようになった (大学入試センター, 2021a)。スタナイン (Angoff, 1984; Clark-Carter, 2005; 前川, 2019) とは, 得点分布を上から (あるいは下から) 順に 4%, 7%, 12%, 17%, 20%, 17%, 12%, 7%, 4% の人数比で 9 つの集団に分割し, どの集団に属するかを表した順位数のことである (表 1)。

表 1 スタナイン (100 点満点テストの場合)

スタナイン	範囲	割合	累積割合
1	$[0, Q_1]$.04	.04
2	$(Q_1, Q_2]$.07	.11
3	$(Q_2, Q_3]$.12	.23
4	$(Q_3, Q_4]$.17	.40
5	$(Q_4, Q_5]$.20	.60
6	$(Q_5, Q_6]$.17	.77
7	$(Q_6, Q_7]$.12	.89
8	$(Q_7, Q_8]$.07	.96
9	$(Q_8, 100]$.04	1.00

表中, $(a, b]$ は, 「テスト得点が a 点より大きく b 点以下である」ことを意味している。 Q_1 は, 第 1 スタナイン分位点であり, 受験者数が 1,000 のとき, 最低得点者から数えて 40 位に位置する点のことである。同様に, 第 8 スタナイン分位点 Q_8 は最低得点者から数えて 960 位に位置する点のことである。8 つの分位点により, 受験者は 9 集団に分割される。そのため, スタナイン」と称される。なお, 9 つの分割割合の根拠は, 標準正規分布を -1.75 から 1.75 まで等間隔に 0.5 刻みの 8 つの点で分割したとき, 各区間を積分した面

積が表 1 の各範囲の割合と同一であるからである。以降, t 番目の累積割合を P_t と表記する。

共通テストの成績は, 過年度成績の利用を認める一部の大学では最大で過去 3 年分の成績を利用することができる (大学入試センター, 2021b)。令和 2 年以前の大学入試センター試験では, スタナインは成績提供されていないので過年度利用はできないが, 令和 3 年以降の共通テストではスタナインも過年度利用できる。

スタナインは, 同一科目の試験において年度間に大きな難易差があるときの比較が便利である。例えば, ある年の国語試験が難しく 160 点としてスタナイン 9 に分類された受験者がいるとする。しかし, 翌年の国語試験が易しかったならば, 前年の 160 点を翌年に過年度利用しようとしても, 翌年の基準では 160 点は高得点でなくスタナイン 8 に分類されるとする。そのとき, 前年にとった 160 点が過小評価されることになる。このとき, スタナイン 9 という成績も持ち越すことができるので, 前年にとった 160 点が適切に評価されることになる。

ところで, スタナイン成績がどれくらい安定しているかについては, スタナイン分位点 $Q_1 \sim Q_8$ (標本スタナイン分位点) が, 離散得点分布の背後に想定される母集団連続得点分布におけるスタナイン分位点 (母スタナイン分位点) をどれだけ精度よく推定できているかに依る。もしも, 標本スタナイン分位点の点推定の精度が低ければ, それらを用いて区切られるスタナインは誤分類を起こしている可能性が高くなる。また, スタナイン成績をどれくらい信用できるかは, それぞれの標本スタナイン分位点の信頼区間にも依存する。信頼区間が小さければ, それだけ標本スタナイン分位点によって区切られたスタナイン成績を信用することができる。信頼区間については, 標準誤差 (standard

error, SE) で評価することができる。

本研究の目的は、テスト得点分布におけるスタナイン分位点の SE を数値実験によって評価することである¹⁾。また、一般にテスト得点は0~100点の離散値をとることが多い。したがって、スタナイン分位点も離散値となる。本来、連続的な確率密度においては、連続値として現れる分位点が離散化されて観測されることによる推定誤差も合わせて検討する。

数値実験で検討する手法は、①移動平均による平滑化と区分線形補完、②正規分布による近似、③カーネル密度推定、④ブートストラップ法の4つである。これらの手法の説明については、第 2.4, 2.5 節を参照されたい。なお、多くの手法がある中で、これら4つの手法を選択した理由は、①と②については簡便さおよび手軽さである。共通テストの分析には、多くの分析者が携わり、短時間でスタナインを含む統計量を計算し公表する必要があるため、簡便でかつ手軽であることは大きな利点である。また、③については、確率密度を推定する上で代表的な手法であること、④については、SE を推定する上での代表的な手法であるからである。

2 方法

2.1 スタナイン

スタナインは、必ずしもよく知られている統計量ではないので、本節で説明する。スタナインによる分類とは「 Q_t ($t = 1, \dots, 8$) 点以下の受験者割合が P_t 未満である受験者をスタナイン t に分類し、そうでなければスタナイン 9 とする」というものである。例えば、60~75 点の受験者にスタナイン 7 が与えられているとき、 $Q_7 = 75$ であり、75 点以下の受験者割合が 89% 未満であることを意味している。これは 76 点以下の受験者割合が 89% 以上であると言ってもよい。仮に 75 点以下の受験者が 80% と大きく 89% を下回るが、76 点以下の受験者が 89.1% とわずかでも 89% を超えるのならば、76 点者はスタナイン 8 に分類される。このように、得点が離散値の場合、同順位の受験者が存在するため、最終的な受験者割合が定義通りの割合分割にならないことがほとんどである。また、仮に、76 点者がたった一人しかおらず、76 点以下の受験者割合がちょうど 89% であるときは、89% 「未満」でなくなるので、この場合、76 点者はスタナイン 8 に分類される。

このようなスタナイン分類基準に従うと、例えば、たった一人しか受験者がいないとき、その得点が何点であっても、その得点は下から 100% 点であるので、

4% 未満、11% 未満、…、96% 未満の基準をことごとく満たさず、この受験者にはスタナイン 9 が与えられることになる。また、もし、受験者数が 2 人で彼らが異なる得点をとったならば、低得点者は下から 50% に位置するため、 P_5 未満を満たすためスタナイン 5 が与えられ、高得点者は下から 100% に位置するため、 P_8 を超えるのでスタナイン 9 が与えられる。

2.2 分位点の離散化による推定誤差

一般に、スタナイン分位点に限らず、標本分位点は漸近的に正規分布することがすでに知られている (e.g., Rao, 1973)。いま、確率変数 X があり、 X の分布関数が $F(x)$ であるとする。また、 F は一次微分可能であるとし、 X の密度関数を $f(x) = F'(x)$ とする。

このとき、 $0 < p < 1$ である p に対し、 F における $100p\%$ 点 Q_p は、

$$\Pr(X \leq Q_p) = F(Q_p) = p \quad (1)$$

を満たす点である。いま、分布関数 F に従う確率変数から n 個のデータ x_1, \dots, x_n を独立に無作為抽出し、これらのデータから $100p\%$ 点 \hat{Q}_p を求めることができる。この標本分位点が漸近的に以下に示す平均と分散を持つ正規分布に従うことが知られている。すなわち、

$$\hat{Q}_p \sim N\left(Q_p, \frac{p(1-p)}{nf(Q_p)^2}\right) \quad (2)$$

である。しかし、離散的なテスト得点分布では、 \hat{Q}_p は離散化されて観測されるので、連続分布における連続的な点であるはずの Q_p の推定に真値とのずれがあるはずである。

2.3 分位点の SE

(2) より、標本分位点 \hat{Q}_p の SE は、

$$SE(\hat{Q}_p) = \sqrt{\frac{p(1-p)}{nf(Q_p)^2}} \quad (3)$$

のように導かれる。スタナイン分位点は、無数の分位点のうち8つの特定の点であるため、(3)式を適用してスタナイン分位点の SE を算出することができる。しかし、このことを行うためには2つの克服すべき課題がある。それらは、

(A) テスト得点分布は離散分布

(B) f の形が未知

である。(A) について、上述したが、我が国におけるテストの多くは取りうる値を 0 点~100 点としている。すなわち、テスト得点は非負の整数値として観測される。したがって、経験分布は離散型である。(3) 式の

標本分位点の SE は、 F が x に関して連続である (x が連続得点である) ことを前提として導出されている。

続いて(B)として、 f (または同じことであるが F) の形が未知である。標本分位点の SE は f の形が既知であってこそ計算可能である。

2.4 連続確率密度の推定

前節で述べた 2 つの課題より、分位点とその SE を計算するためには、まず、観測された離散得点分布から連続分布 f を推定する必要がある。本研究では、以下の 3 つの方法を用いる。

- ① 3 項移動平均による平滑化と区分線形補完
- ② 正規分布による近似
- ③ カーネル密度推定

①では、まず、得られた離散経験分布を移動平均 (moving average, MA) によって平滑化する。いま、離散得点分布の累積相対度数分布を

$$F = \{R_0, R_1, \dots, R_{100}\} \quad (4)$$

とする。このとき、3 項 MA によって平滑化後の分布 F_M を以下のように得る。

$$F_M = \{Y_0, Y_1, \dots, Y_{100}\}$$

$$Y_t = \begin{cases} (R_t + R_{t+1})/2 & (t = 0) \\ (R_{t-1} + R_t + R_{t+1})/3 & (t = 1, \dots, 99) \\ (R_{t-1} + R_t)/2 & (t = 100) \end{cases} \quad (5)$$

次に、 F_M に対して区分線形補完を行う。つまり、得点 t ($= 0, \dots, 99$) と $t + 1$ の区間を連続的に

$$F_M(x) = (Y_{t+1} - Y_t)(x - t) + Y_t \quad (6)$$

のように線形補完する。この区分線形関数を微分し、全区間を積分した結果が 1.0 になるようにした

$$f_M(x) = \frac{F'_M(x)}{F_M(100) - F_M(0)} \quad (7)$$

を f とする。

続いて、②の正規分布による方法は、経験分布の平均と分散を、それぞれ \bar{x} と s^2 とすると、それらを正規分布の平均と分散に指定した正規密度関数 $f_N(x|\bar{x}, s^2)$ を用いて、

$$f_N(x) = \frac{f_N(x|\bar{x}, s^2)}{F_N(100|\bar{x}, s^2) - F_N(0|\bar{x}, s^2)} \quad (8)$$

を f とする。ここで、 $F_N(\cdot|\bar{x}, s^2)$ は、 $f_N(\cdot|\bar{x}, s^2)$ の累積分布関数であり、 f_N は $0 \leq x \leq 100$ の範囲において、面積が 1.0 になるように調整されている。

さらに、③のカーネル密度推定 (kernel density estimation) による方法は、多様な変形があるが、本研究では、分散が等しい C 個の正規分布が混合して

いると仮定し、共通分散 σ^2 と C 個の混合比率 $(\pi_1, \dots, \pi_C; \sum_C \pi_C = 1)$ を推定することで、確率密度 f を得る方法 (e.g., Ferguson, 1973; Escobar & West, 1995; Muller, Erkanli, & West, 1996) を用いる。混合分布の数 C を推定対象とすることもできるが、本研究では、 $C = 101$ とし、 $c + 1$ 番目の分布の平均を c とする。したがって、この方法によって推定される分布は

$$f_K(x) = \frac{\sum_{c=0}^{100} \pi_c f_N(x|c, \sigma^2)}{\sum_{c=0}^{100} \pi_c F_N(100|c, \sigma^2) - \sum_{c=0}^{100} \pi_c F_N(0|c, \sigma^2)} \quad (9)$$

となる。 f_K もまた面積が 1.0 に調整されている。

図 1 は、①~③の方法による確率密度の推定の結果を示したものである。網掛け(と細い実線)の分布 f_{Obs} は、経験分布である。1 点刻みの区間でプロットしているため、棘状化(ギザギザ)していることがわかる。

経験分布 f_{Obs} に対して、方法①の移動平均によって平滑化された分布 f_M は太い実線で示されている。棘状化が緩和している。今回は、3 項移動平均を用いているが、5 項移動平均などのより強い平滑化を用いれば、より滑らかな分布が得られる。

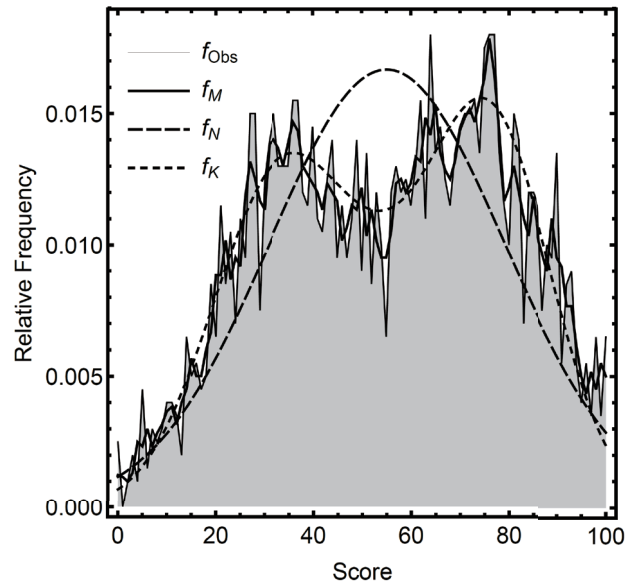


図 1 3 つの方法による確率密度の推定

また、長破線は、方法③の正規近似密度 f_N である。経験分布が正規分布に近くないとき、正規分布は良い近似とならない。さらに、短破線は、カーネル密度推定された分布 f_K である。経験分布のギザギザの中央を滑らかに通るように推定されている。

確率密度が推定されたら、(1)式によりスタナイン分位点を算出でき、(3)式により SE を推定することができる。なお、本節で述べた方法は、分位点を評価する際の誤差に連続

分布の誤推定も含まれる。

2.5 ブートストラップ法

2.2 節で述べたように、分位点と SE を評価するには、連続分布 f を推定する必要がある。しかし、ブートストラップ法 (Efron, 1979; Efron & Tibshirani, 1994; 汪・田栗, 1996; 汪・桜井・金, 2011) を用いれば、 f を推定することなく分位点と SE を求めることができる。例えば、中央値も分位点の 1 つ 50% 点であるので、中央値とその SE を、それぞれ、(1) 式と (3) 式を用いて評価することができる。しかし、中央値とその SE をブートストラップ法で推定することは、ブートストラップ法の入門的な課題としてよく紹介されている。

ブートストラップ法は、 n 個のデータがあるとき、 n 個のデータから重複を許しながら n 個のデータをリサンプリングし、興味のある統計量を計算する。この操作を B 回繰り返すと、 B 個の推定値が得られる。そして、 B 個の推定値の平均を点推定値、SD を SE とする。特に、解析的に求めるのが困難な統計量の SE を評価することに優れた手法である。

なお、リサンプリングによる方法は、他にもジャックナイフ法 (Quenouille, 1956) がよく知られているが、中央値やトリム平均などの SE の推定は、ブートストラップ法の方が精度が高い (Efron, 1982)。本研究では、方法④として、 $B = 1000$ としたブートストラップ法を用いてスタナイン分位点と SE を評価する。

3 数値実験

3.1 データ発生

本研究では、真の分布として、3 つの分布を仮定する (図 2)。それらは、

$$\begin{aligned} f_1(x) &\propto f_N(x|65, 20^2) \\ f_2(x) &\propto f_N(x|100, 50^2) \\ f_3(x) &\propto .45f_N(x|35, 15^2) + .55f_N(x|75, 15^2) \end{aligned} \quad (10)$$

と設定した。

f_1 は平均が 65、SD が 20 の切断正規分布である。共通テストでは、平均が 65 付近で、SD が 20 付近である科目が多いので、この分布を設定した。 f_2 は、モードが 100 であるような切断正規分布である。この分布は、モード (平均と中央値も) が 100 になっている分布であり、非常に易しい科目でこのような分布がまれに観測されることがある。このような形状の分布でもスタナインの振る舞いを確認する必要があるため設定した。また、 f_3 は、2 つの正規分布が混合した切断混合正規分布である。共通テストでは、現浪など、学力

差が大きい集団が混在するが、時としてその差が目立つことがあるので、このような形状の分布を設定した。

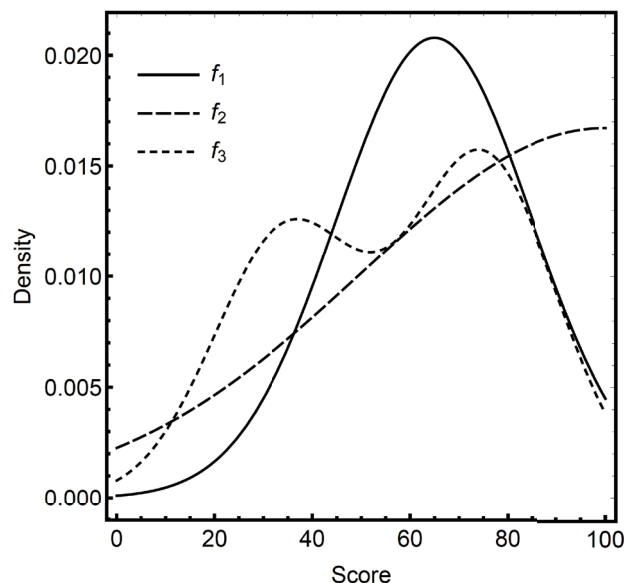


図 2 データ発生分布

確率密度が既知のとき、スタナイン分位点とその SE の理論値が分かる (表 2)。分位点の理論値は (1) 式を満たす点を求めればよい。また、SE の理論値は、(3) 式で求まるが、正確には標本サイズ n が未知のもとでは求まらないので、表 2 には、標本サイズの平方根で除する前の SE の理論値を掲載している。

表 2 スタナイン分位点と SE の理論値

	分位点			$\sqrt{n}SE$		
	f_1	f_2	f_3	f_1	f_2	f_3
Q_1	29.74	13.51	15.59	44.58	52.32	37.42
Q_2	40.05	28.11	24.94	32.77	52.61	32.21
Q_3	49.64	44.27	35.40	27.18	46.85	33.52
Q_4	59.12	60.31	49.51	24.60	40.16	43.76
Q_5	68.84	75.08	66.03	24.00	33.18	34.33
Q_6	77.82	86.06	77.11	24.86	26.17	27.25
Q_7	86.11	93.40	85.69	26.27	18.88	26.17
Q_8	93.31	97.61	92.85	25.67	11.73	25.87

このような特徴を持つ 3 つの分布から $[0, 100]$ の範囲でサンプリングし、四捨五入して離散化する。標本サイズは、500、1000、2000 の 3 パターンとした。共通テストには、受験者数が 10 万以上の大受験者科目があるが、そのような大受験者数のもとでは分位点は正しく求まっているので問題ないが、共通テストには受験者数が 1000 に満たない少数受験者科目もあり、そのような

少数受験者科目において分位点の精度を調べる必要があるため、このような標本サイズの設定とした。

本数値実験計画は、「3 パタンの真の分布 (f_1, f_2, f_3) のもとで、3 パタンの標本サイズ (500, 1000, 2000) で発生させた離散型データに対して、①~④の方法を用いて、スタナイン分位点とその SE を評価する」と要約できる。また、各(3×3×4)条件での繰り返しは 100 とした。スタナイン分位点と SE の真値の再現性の指標としては、平均絶対差 (mean absolute difference, MAD) を用いる²⁾。MAD は、

$$MAD = \frac{\sum_{r=1}^{100} |\tau_r - \tau|}{100} \quad (11)$$

と定義される。ここで、 τ は真値である。本研究では、8 つのスタナイン分位点とその SE (Q_t と $SE(Q_t)$; $t = 1, \dots, 8$) が相当する。また、 τ_r は、 r 回目の繰り返しにおける τ の推定値である。MAD が 0 に近いほど、真値に対する絶対的な距離に関して、平均的な再現性に優れていることを表す。

3.2 結果

スタナイン分位点の MAD を図 3 に示す。図中のマーカーの D, M, N, K, B は、それぞれ、離散化 (discretized) による影響、①移動平均、②正規近似、③カーネル密度推定、④ブートストラップ法を意味している。また、それらマーカーに付随している 5, 10, 20 は、それぞれ、標本サイズが 500, 1000, 2000 の結果を表している。例えば、B5 であれば、標本サイズが 500 におけるブートストラップ法の MAD を表している。また、上・中・下のグラフは、それぞれ、真の分布が f_1, f_2, f_3 (図 2 参照) のときの結果である。

図 3 より、標本サイズが大きくなるほど、真値の再現性が向上していくことが分かる。また、各標本サイズにおいて、もっとも平均的に ($Q_1 \sim Q_8$ を通じて) MAD が優れていた方法のマーカーを大きく表示し、実線で結んでいる。例えば、真の分布が f_1 のとき (図 3 上図)、標本サイズが 500 のもとでは、B5 のマーカーが大きく表示され実線で結ばれている。これは、この条件では、ブートストラップ法が最もスタナイン分位点の再現性が、 $Q_1 \sim Q_8$ を通じて平均的に優れていたことを意味している。図 3 の上中下の図より、ブートストラップ法が、スタナイン分位点の再現性に最も優れていたことを表している。

同様に、図 4 は、スタナイン分位点の SE の再現性を示したものである。1 つの例外を除く全ての条件においてもカーネル密度推定を用いて連続分布を推定し、

SE を評価する方法が最も再現性が高かったことを示している。

4 結論

離散化されたテスト得点分布でスタナイン分位点の点推定値を算出するとき、経験分布から単純にスタナイン分位点を求めると、点推定値に誤差を伴うことが避けられないことがあると分かった。数値実験の結果、本研究で用いた 4 つの方法の中では、真のスタナイン分位点に対して、絶対距離において最も平均的に小さく迫ることができるのは (すなわち MAD の観点から)、ブートストラップ法であることが分かった。

一方で、スタナイン成績の確からしさは、スタナイン分位点を区間推定したときの区間幅の小ささに依存するので、そのときスタナイン分位点の SE を評価することが肝要である。MAD の観点からは、スタナイン分位点の SE を最も精度よく推定したのは、本研究で用いた 4 つの方法の中ではカーネル密度推定で連続分布を推定した後に(3)式を用いて SE を評価する方法だった。

本研究の限界は、本研究では真の分布の設定が 3 つであり、さまざまな分布の形状について調べつくしたわけではないことである。また、カーネル密度推定とブートストラップ法に関しては、様々な細かい設定 (チューニングパラメタなど) の下で調べつくしたわけではなく、設定の仕方によっては、本研究の結果よりも良い結果を報告しうる設定は十分ありうる。

また、数値実験における再現性の指標については、点推定値の正負の方向性を考慮した「偏り (バイアス)」の検討も大事であるが、本研究においては報告しなかった。分位点の位置によって偏りに一貫した結果を見出せなかったからである。偏りは、データを発生させるときの得点分布 (図 2 参照) の形に強く影響していると思われるため、この点については、さまざまな得点分布の形を検討する中で改めて議論する機会を持ちたい。

さらに、今後の課題は、本研究で用いた方法が 4 つであり、より高精度に真値を再現する他の方法を探す、あるいは開発することである。その 1 つの方向性としては、連続分布を仮定せず、離散分布の分位点の信頼区間を直接的に評価する方法であろう。

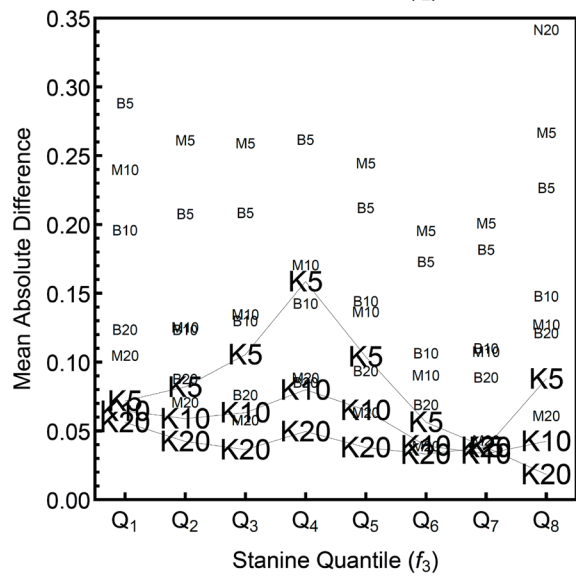
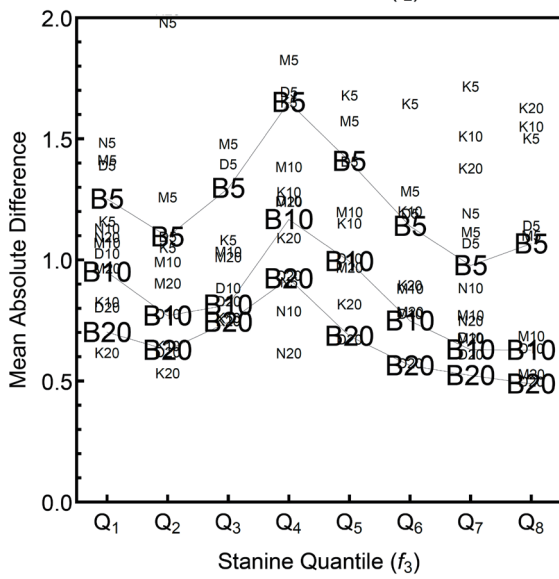
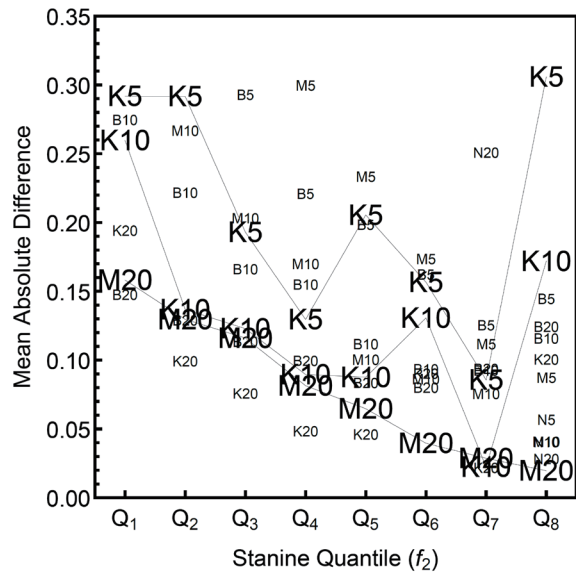
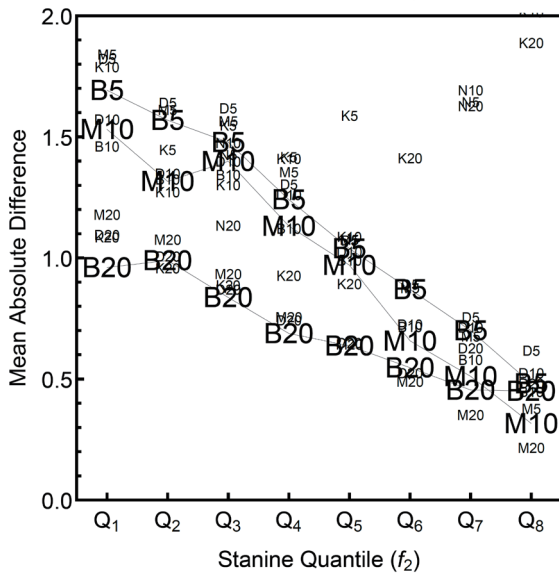
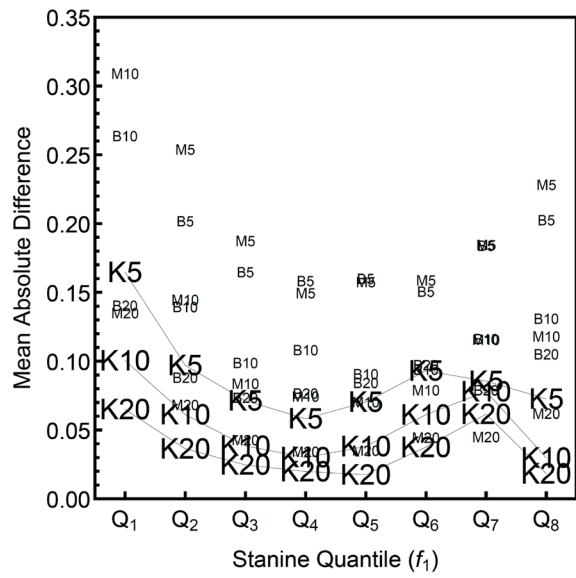
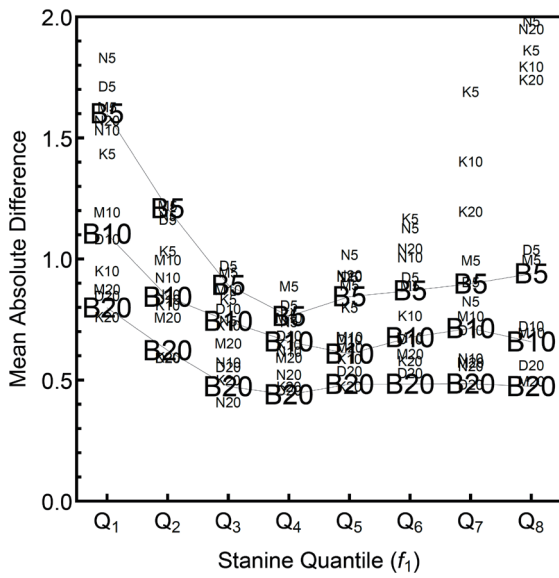


図3 スタナイン分位点の再現性

図4 スタナイン分位点のSEの再現性

注

- 1) 本論文の見解は、著者らの個人的見解であり、所属組織の公式見解ではありません。
- 2) 再現性の指標として、絶対差平方平方根 (root mean squared difference, RMSD) もよく用いられるが、MAD の優劣づけとはほぼ一致したので省略する。

謝辞

本研究は、大学入試センター理事長裁量経費研究「令和 7 年度共通テストに対応した得点調整方法と成績提供法の検討」の助成を受けたものである。

参考文献

- Angoff, W. H. (1984). Scales, norms, and equivalent scores. Educational Testing Service. (Reprint of chapter in R. L. Thorndike (Ed.) (1971) *Educational Measurement (2nd Ed.)*. American Council on Education.
- Clark - Carter, D. (2005). Stanine scores. Brian S. Everitt & David C. Howell (Eds). *Encyclopedia of statistics in behavioral science*. Volume 4, pp. 1893. <https://doi.org/10.1002/0470013192.bsa640>
- 大学入試センター (2021a). 「令和 3 年度大学入学共通テスト段階表示換算表」大学入試センター (dnc.ac.jp/albums/abm.php?f=abm00040264.pdf&n=段階表示換算表.pdf) (2021年11月25日).
- 大学入試センター (2021b). 「令和4年度大学入学選抜に係る大学入学共通テスト受験案内」大学入試センター ([https://www.dnc.ac.jp/albums/abm.php?f=abm00040523.pdf&n=令和4年度受験案内\(50MB\).pdf](https://www.dnc.ac.jp/albums/abm.php?f=abm00040523.pdf&n=令和4年度受験案内(50MB).pdf)) (2021年11月25日).
- Efron, B. (1979). Bootstrap methods: Another look at the Jackknife. *The Annals of Statistics*, 7, 1–26.
- Efron, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans*. Society of Industrial and Applied Mathematics CBMS-NSF Monographs, 38.
- Efron, B. & Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. Chapman & Hall.
- Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric Problems. *The Annals of Statistics*, 1, 209–230.
- Escobar, M. D. & West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of American Statistical Association*, 90, 577–588.
- 前川眞一 (2019). 得点の段階表示とStanine. 大学入試センター研究開発部リサーチノート RN-19-01.
- Muller, P., Erkanli, A., & West, M. (1996). Bayesian curve fitting using multivariate normal mixtures, *Biometrika*, 83, 67–79.
- Quenouille, M. (1956). Notes on bias in estimation. *Biometrika*, 43, 353–360.

Rao, C. R. (1973) *Linear Statistical Inference and its Applications (2nd Ed.)*. Wiley.

汪金芳・桜井裕仁・金明哲 (2011). 『ブートストラップ入門』共立出版.

汪金芳・田栗正章 (1996). 「ブートストラップ法」『統計数理』44, 3–18.