

素点を用いた DTF検出法における集団間の平均値差補正に関する研究 ー素点に基づいたシミュレーション研究を通じてー

○熊谷龍一（東北大学）

坂本佑太郎（株式会社 リクルートマネジメントソリューションズ）

Back Ground

【DTF】（Differential Test Functioning）：DIFをテスト全体に拡張したもの。弱いDIFが集まることで、テスト得点でみた場合に集団間で差があるかを検討できる。

【Index S】：Sakamoto & Kumagai (2025) で提案された、「素点」に基づくDTFの指標。2集団間の平均値差からIndex S の値を減算することで、DTFの影響を補正することも提案。

Index S

下位集団数が2，2値型データの場合のIndex S によるDTF検出手順は以下の通り。

【Step 1】従来のDIF分析を行い，DIFが生じていない項目（アンカー項目）を特定する。

【Step 2】アンカー項目の合計点を算出し，その得点ごとに層分けを行う。各層でテスト全体の群間平均点差に各層の相対度数をかけた d_L を算出する。

$$d_L = (S_{RL} - S_{FL}) \times W_L$$

ここで L は層， S_{RL} ， S_{LF} は層 L における下位集団 R と F のテスト全体の平均点， W_L は全受験者に対する層 L の人数比率である。

【Step 3】以下のIndex S をDTFの大きさを表す指標とする。

$$Index S = \sum_L d_L$$

この指標は，20点満点（素点）のテストであれば「20点満点」の尺度上でDTFの影響がどの程度であるかを表すことができる。またDTFが生じている場合に，2つの集団間平均値差からIndex S の数値を引くことで，DTFの影響を取り除いた平均値差（平均値差の補正）を示すことができる。

Problem & Aim

Sakamoto & Kumagai (2025) では，Index S の検証をIRTモデルを利用したシミュレーション研究で行っている。IRTモデルでは，真値の設定として θ が用いられるため，素点での平均値差の補正程度に関する検証を行うことが難しい。そこで本研究では，素点に基づいたシミュレーション研究を行うことで，平均値差の補正について検証を行う。

Simulation Steps

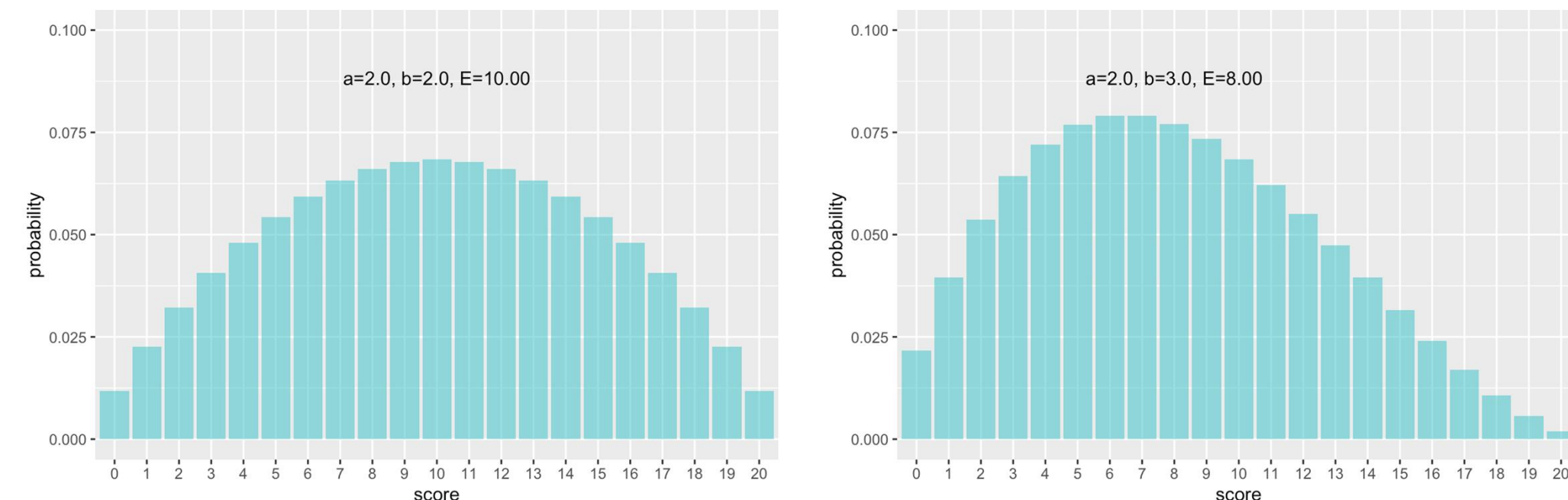


Figure 1. ベータ二項分布

Step 1. 素点（合計点）の母集団分布の設定

2つの集団の素点において，ベータ二項分布を母集団分布と設定する。ベータ二項分布は a , b の2つの母数によりその形状と期待値・分散を設定できる。この時点で2つの集団の「①母集団平均値（期待値）の差」を計算することができる。

Step 2. 素点（合計点）の発生

ベータ二項分布に従う乱数により，各集団の素点データを発生させる（例えば，各集団500個）。このデータにより，「②標本の集団平均値差」を計算することができる。

Step 3. 正誤（0-1）データの発生

各受験者において項目数分の0ベクトルを作成し，その中から重みづけ無作為抽出により，素点（合計点）の数だけ1に変換する。その際の重みづけとして，項目1から4は(0.2, 0.4, 0.6, 0.8)とした。項目5以降はその重みを繰り返して利用する。このため全項目数は4の倍数とする。

Step 4. DIFを利用したテストバイアスの混入

全項目のうちバイアスがない項目（アンカー項目）とバイアスを受ける項目（DIF項目）を設定し，集団2のDIF項目において正答率を計算する。その後，正答率を $n\%$ （例えば10%）下げるように，データの1を0に無作為に変換する。この処理を行ったデータ行列に対し，2集団の平均値差を求めることで，「③DTFの影響が混入した平均値差」を計算することができる。

Step 5. index S の計算

Sakamoto & Kumagai (2025)の手続きに従いindex S を計算し，「③DTFの影響が混入した平均値差」からindex S を引くことで「④補正後の平均値差」を計算することができる。なおindex S の計算において，本研究ではアンカー項目は既知であるとした。

Step 6. 繰り返し計算

上記2.～5.を500回繰り返し，②，③，④の3種類の平均値差を500個算出する。

Result

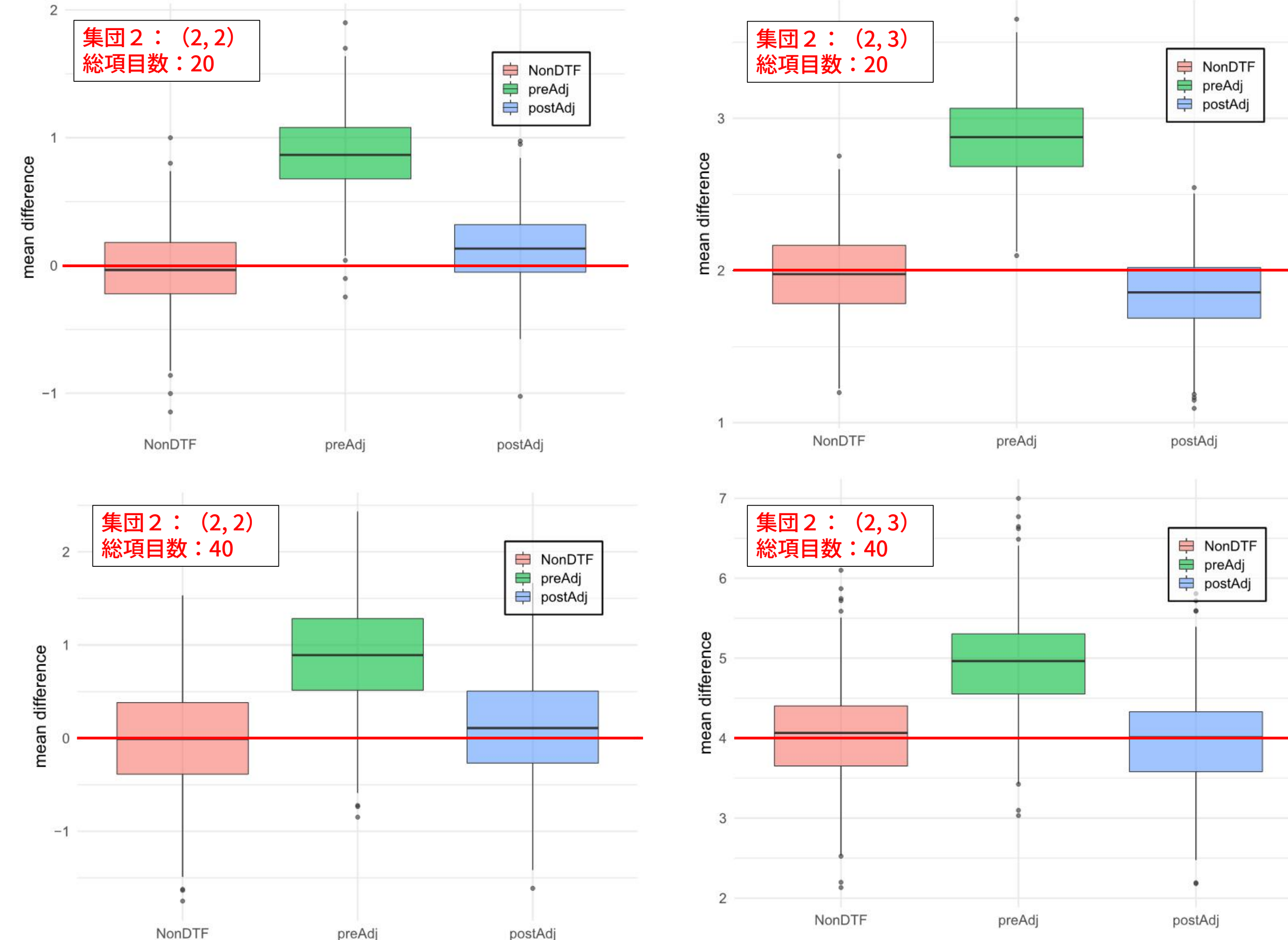
シミュレーションにおける設定値をTable 1に示す。

Table 1. シミュレーション設定値

受験者数（素点の数）	集団 1：500， 集団 2：500
ベータ二項分布母数	集団 1：(2, 2) 集団 2：(2, 2) or (2, 3)
総項目数	20 or 40
DIF項目数	8
DIFの影響	集団 2 の正答率を10%低下

シミュレーション手続きにおける②～④の平均値差（500回分）の箱ひげ図をFigure 1に示す。なお，図中のNonDTF が②，preAdjが③，postAdj が④の平均値差であり，赤線は①の平均値差（理論値）である。

Figure 2. 3種類の平均値差の箱ひげ図



Discussion

どの設定条件においても，DIF項目によるテストバイアスの影響により，2集団間の平均値差にバイアスが生じた（図中，preAdj）が，index S による補正が適切に理論値方向に機能していることが確認された（図中，postAdj）。

ただし，総項目数に対するDIF項目の割合が大きい場合（総項目数20に対して，DIF項目数8）においては，完全に補正しきれているわけではないことも確認された。

2群の平均値差の比較は，教育心理学関連の研究においても，非常によく用いられる分析である。多くの研究においては，2群が受検しているテスト・質問紙にはDIF・DTFの影響はなく，2群の平均値差は，その研究が注目としている要因によるものであることが前提条件となっている。

index S 自体の計算は非常に簡便に行われるため（ただしDIF項目の選定は様々な手法がある），2群の平均値差を算出する際には，一度DTFの影響を調べたうえ，必要に応じて補正などをする必要があると言えよう。

文献

Sakamoto, Y., & Kumagai, R. (2025). A new method for detecting DTF using sum scores. 2025 International Meeting of the Psychometric Society.

This work was supported by JSPS KAKENHI Grant Number JP25K06765.