

素点を用いたDTF検出法における標準化指標の提案

坂本 佑太郎¹・熊谷 龍一²

(¹株式会社リクルートマネジメントソリューションズ)

(²東北大学大学院教育学研究科)

問題と目的

- Differential test functioning(DTF)
 - DIF分析を行う際には、項目レベルの検証だけではなく、テスト(尺度)全体での差異である特異テスト機能(Differential test functioning, DTF)も同時に検証する必要がある(Chalmers, Counsell & Flora, 2016 ; Temel, 2023)
- 素点を使ったDTF検出法
 - 昨今、素点合計点の実用性について再評価 (Sijtsma et al., 2024).
 - Sakamoto and Kumagai (2025)は素点を使ったDTF検出法を提案
 - この方法は、たとえば20点満点のテスト同士の得点を比較したい時、「DTFの影響がX点分ある」という、素点尺度上でその影響度合いを捉えられる利点がある
 - 一方で、標準化指標は提案されていない
- 本研究の目的
 - 素点を使ったDTFの検出指標“Index S”を標準化指標へと拡張すること

提案手法

- 下位集団数が2および2値型データの場合のDTF検出手順は以下の通り
- STEP3まではSakamoto and Kumagai (2025)で提案しており、STEP4が本研究での提案となる：

Step 1 DIFがない項目をアンカー項目として特定する

Step 2 アンカー項目の合計点を算出した上で、その得点ごとに層分けを行い、各層内で群間のテスト全体の平均点差に重みを乗じたもの (d_L) を算出する

$$d_L = (\bar{S}_{RL} - \bar{S}_{FL}) \times W_L.$$

ここでは L は層を、 $\bar{S}_{RL}$ と $\bar{S}_{FL}$ は、層 L における下位集団 R と F のテスト全体での平均点、 W_L は全受検者に対する層 L の人数比率を表す
なお、当該層の人数が少ない場合(たとえば5人より少ない場合)は、隣接する層を統合する

Step3 以下のIndex SをDTFの大きさを表す指標とする

$$\text{Index } S = \sum_L d_L.$$

この指標は、たとえば20点満点のテストであれば、「20点満点という尺度」上でDTFの影響がどの程度の大きさかを表すことができるその点において、既存のDTF指標 (Shealy & Stout, 1993) とは異なる

Step4 最後にIndex S を全受検者のプールした標準偏差で除すことで、標準化された効果量として解釈できる指標となる

$$\text{Index } S_{std} = \frac{\text{Index } S}{SD_{pooled}}$$

シミュレーション実験

下位集団数が2(集団RとF)および2値型データにおいて、以下の条件でシミュレーション実験(500回)を実施した

Table1シミュレーション実験の条件一覧

Design factor	Value
Respondents per group	$N_R = N_F = 500$
Replications per condition	500
Test length	20, 40
Proportion of DIF	10%, 30%, 50%
Type of DIF	uniform DIF and non-uniform DIF
Direction of DIF	“all” and “half & half”
Distribution of latent trait	$\theta_R \sim N(0,1)$ and $\theta_F \sim N(0,1)$ $\theta_R \sim N(0,1)$ and $\theta_F \sim N(-0.5,1)$

結果

- Index S_stdとIRTを使ったDTF検出法における標準化指標であるETSSD(expected test score standardized difference) (Meade, 2010)の相関を確認した (Figures 1-4).

Figure 1 Index S_std and ETSSDの関係 (Test length = 20, Prop of DIF = 10%, Type of DIF = uniform DIF, Direction of DIF = “all”, $\theta_R \sim N(0,1)$ and $\theta_F \sim N(0,1)$)

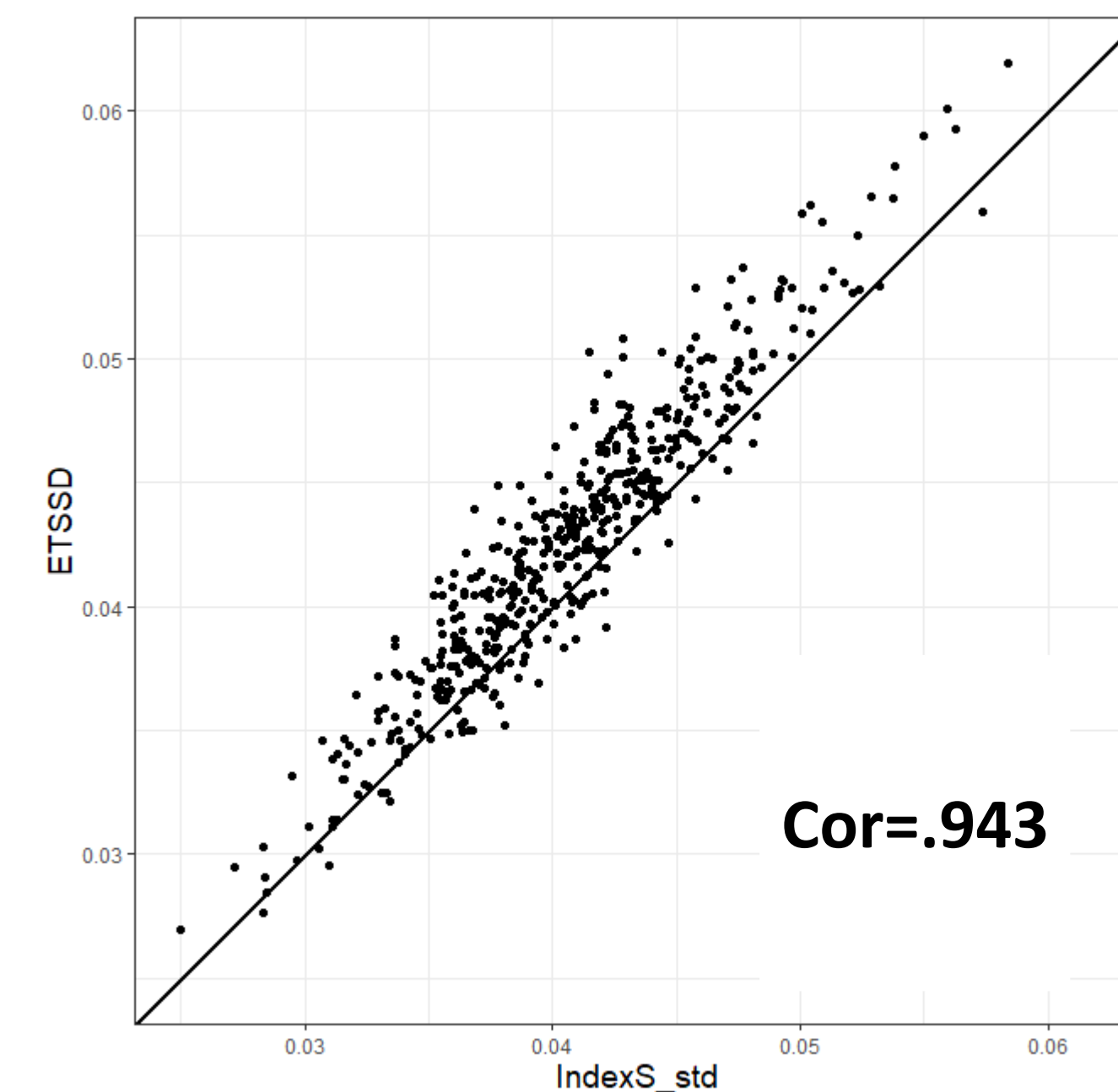


Figure 2 Index S_std and ETSSDの関係 (Test length = 40, Prop of DIF = 30%, Type of DIF = non-uniform DIF, Direction of DIF = “half & half”, $\theta_R \sim N(0,1)$ and $\theta_F \sim N(0,1)$)

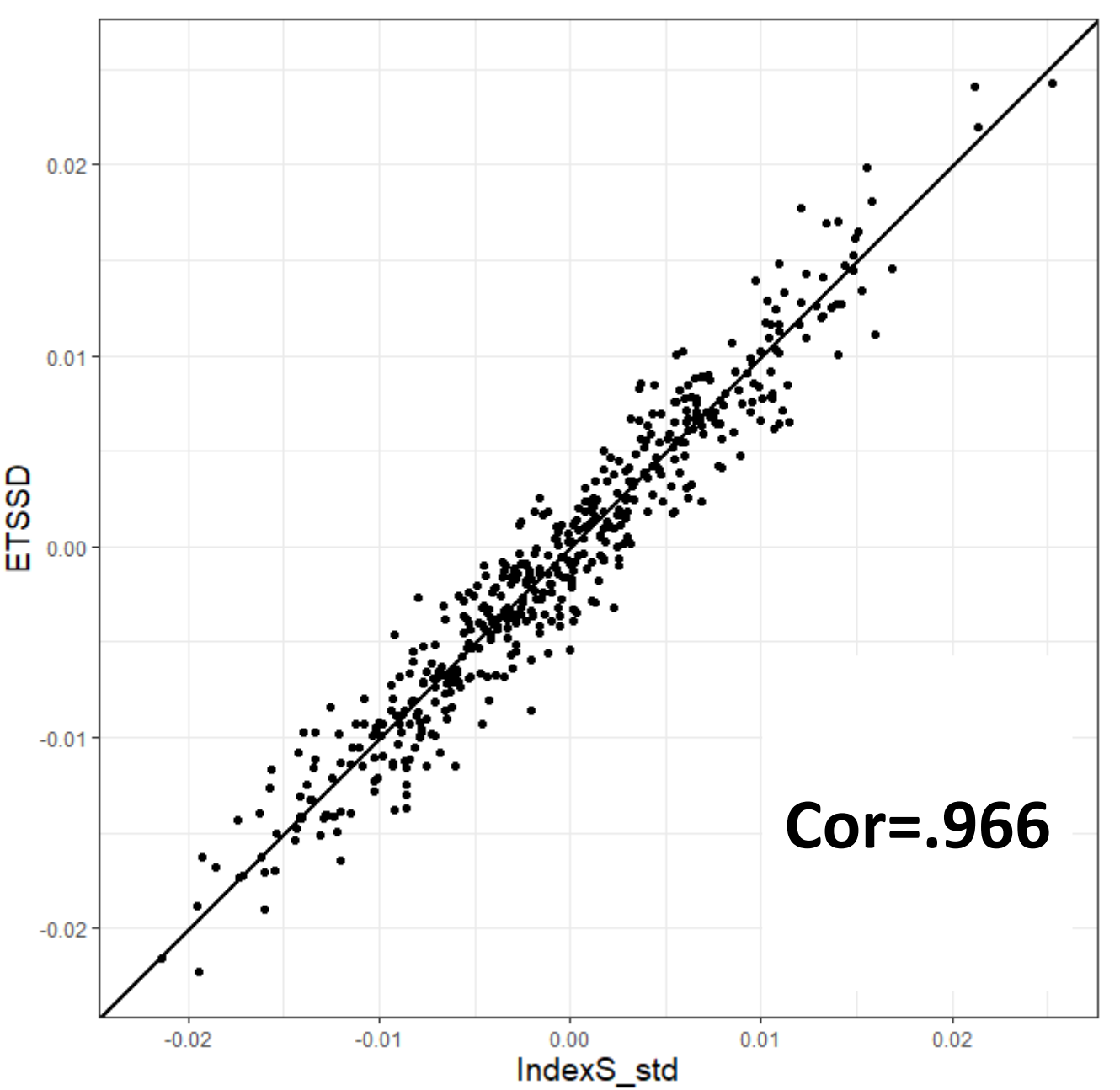


Figure 3 Index S_std and ETSSDの関係 (Test length = 20, Prop of DIF = 10%, Type of DIF = uniform DIF, Direction of DIF = “all”, $\theta_R \sim N(-0.5,1)$ and $\theta_F \sim N(0,1)$)

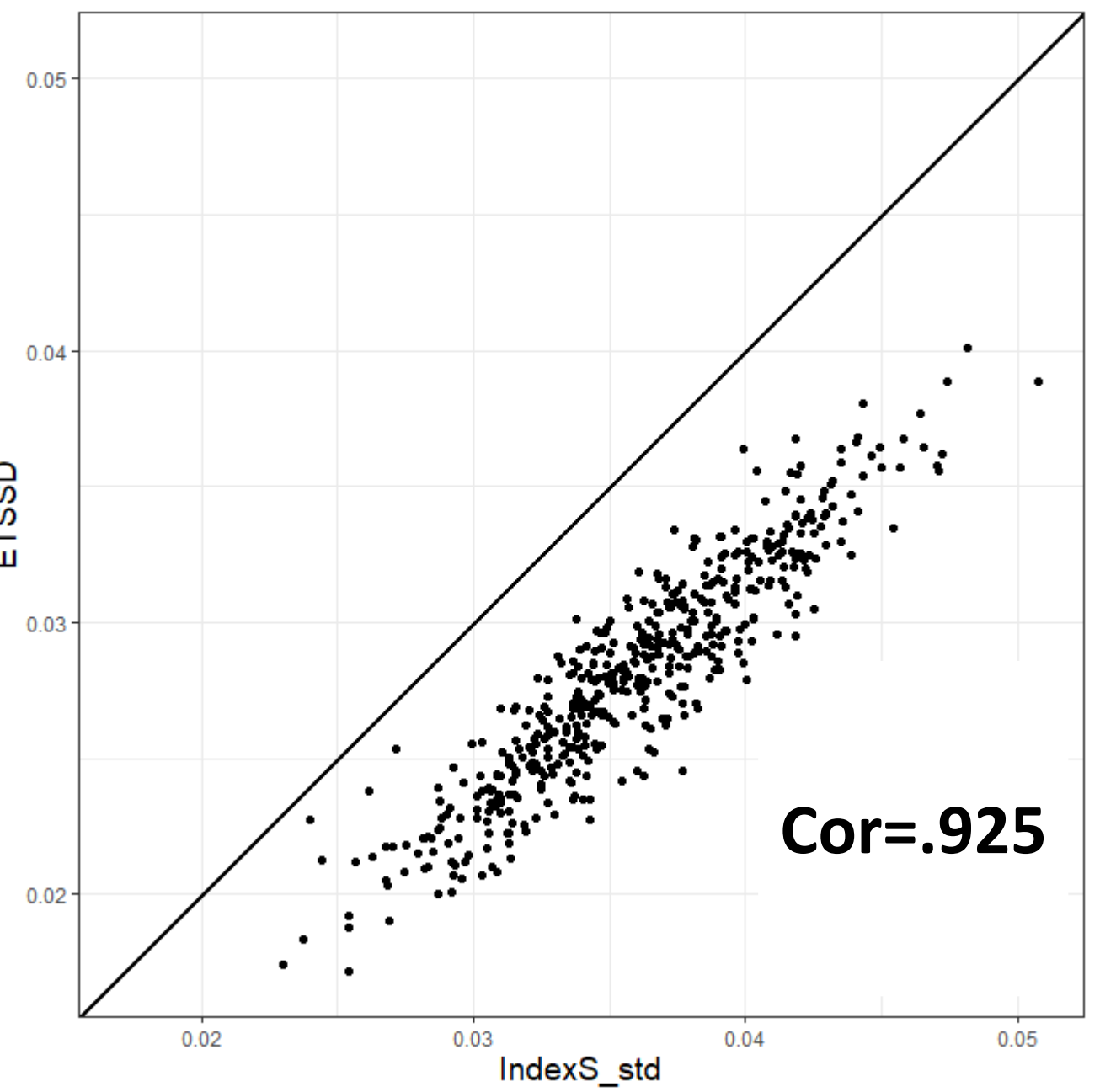
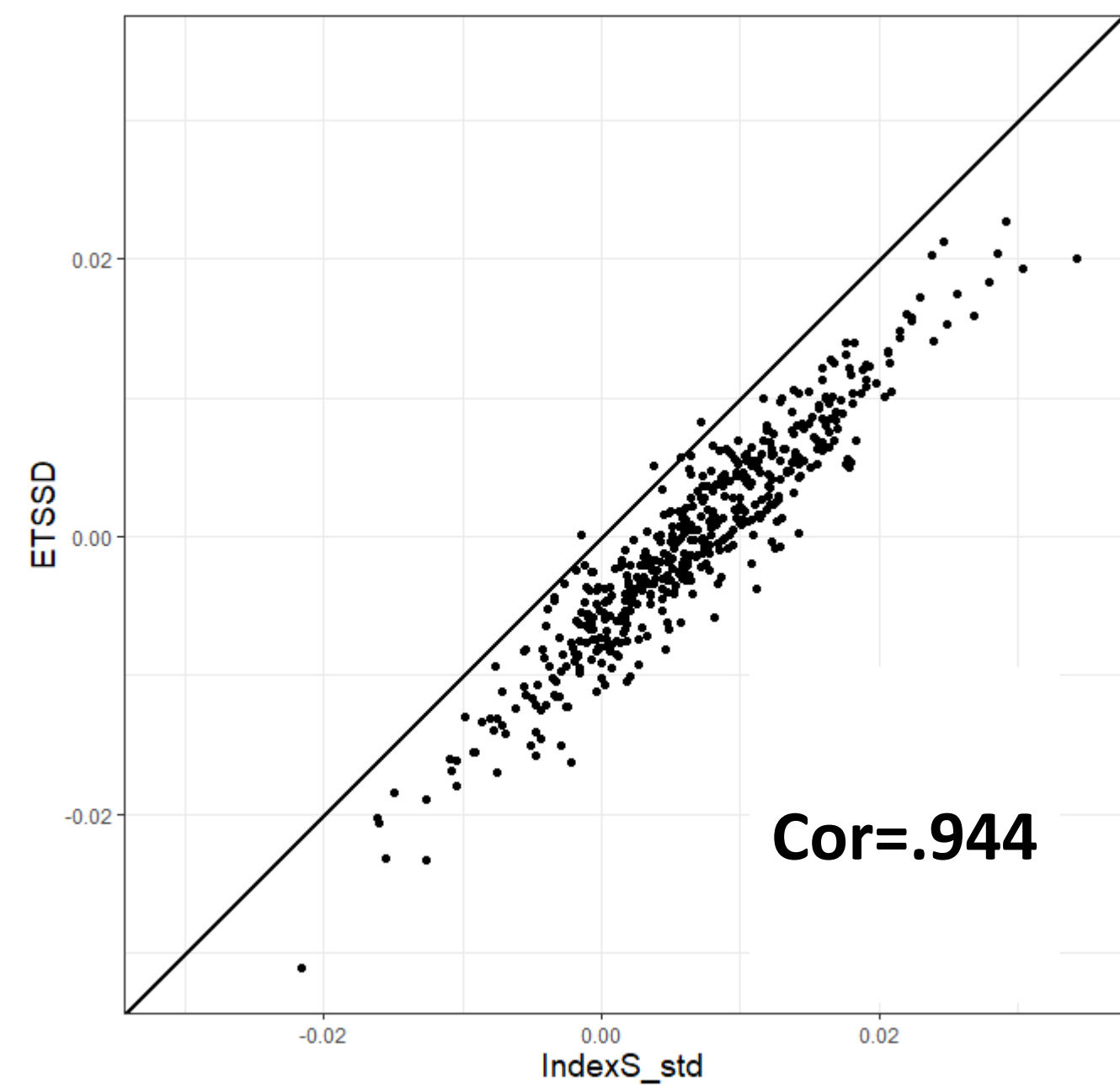


Figure 4 Index S_std and ETSSDの関係 (Test length = 40, Prop of DIF = 30%, Type of DIF = non-uniform DIF, Direction of DIF = “half & half”, $\theta_R \sim N(-0.5,1)$ and $\theta_F \sim N(0,1)$)



- 上記に加え、他の条件も含めてIndexS_stdとETSSDの相関は .7～.8以上
- 潜在特性尺度値の分布を異なるように設定したとき (Figure3, 4), IndexS_stdとETSSDの挙動が異なる
- ✓ ETSSD: 潜在特性尺度値の分布が異なるだけにも関わらず、DTFの大きさが変化している可能性 (Figure1→3が顕著)
- ✓ IndexS_std: 若干の変化はあるものの、DTFの大きさ自体は潜在特性尺度値の分布が異なる場合でもそれほど変化していない可能性 (Figure1→3, 2→4)

考察と限界

- 上記の結果から、IndexS_stdは素点に基づいた方法ながらもIRTを使った指標と同等の性能を持つ可能性が示唆された
- ただし、IRTを用いた方法との挙動が異なる部分があり、今後精査が必要
- 具体的な今後の課題としては、以下がある：
 - IndexS_stdとETSSDとの挙動精査
 - 「真の素点平均値差」の復元の試み
 - d_L 算出の際に $\bar{S}_{RL} - \bar{S}_{FL}$ の符号が逆転する場合の対応
 - 3群以上を比較する場合への一般化 等