

基礎の基礎

45分で学ぶ テスト作りの流れ
項目反応理論

項目反応理論 —おはなしをひと握りだけ—

1. テスト理論
2. 古典的テスト理論
3. 項目反応理論

テストとテスト理論

- テスト(検査):学力や性格などの心理的な特性を測定する用具
- テスト理論:テストの作成方法, 実施方法, 採点方法, 解答の分析方法などに関する知識体系
- 良いテストの条件 ……みなさんおなじみ
 - ✎ 適切に心理的な特性を測定している = 妥当性
 - ✎ 信頼のおける安定した得点を算出できる = 信頼性

古典的テスト理論の問題点例

□ 平行検査法 (parallel-forms estimation approach)

平行検査の分布, 真値, 期待値, 全ての組合せの共分散は等しいとする。

→ 厳密に条件を満たす平行検査を複数作ることは難しい。

□ 再検査法 (test-retest estimation approach)

同一被験者の同一質問に対する2つの反応が独立したもので, 測定誤差がランダムであると考ええる。

→ 被験者が学習能力のある人間である限り無理がある。

古典的テスト理論の問題点例

□ 素点のテスト依存性

正答数(素点)がテスト依存である限り, 素点そのものを独立した学力の測定値として用いることに限界がある。

□ 項目困難度の標本依存性

「この項目を解いた受験生のうち何人が正解したか」という通過率を出す。しかし, 項目困難度を表す統計値としての平均項目得点あるいは通過率が標本依存である。

→通過率の低い難しい項目も, 標本によっては通過率の高い易しい項目になるかもしれない。

たとえば なんかアラサー向けテスト

マスオさん(28歳)が
70点とった

	冊子A	冊子B
『サザエさん』 グループ	解答	
『アラレちゃん』 グループ		解答

則巻博士(28歳)が
60点とった

則巻博士負けてる！！ 天才科学者なんじゃ！？
そうです。実はマスオさん，早稲田大商学部出身。

てことでは終われない

- 冊子A＝基礎問題，冊子B＝応用問題 かもしれず，マスオさんが則巻博士より高い特性であるとは言えない。
→得点(素点)は測定したい特性だけでなく，冊子間の難易差も反映している。
- 偏差値：マスオさんが60，則巻博士が50だったとしても，アラレちゃんグループの方が高得点集団かもしれない。
→偏差値は，集団間に何らかの特性の差異があっても，それを無視している。
- 項目の通過率：冊子Aの問1が0.5，冊子Bの問1が0.6だったとしても，上記と同じ理由から，冊子Aの問題1が冊子Bの問題1より難しいとは言えない。
→通過率は，冊子の難易度だけでなく集団の特性差も反映している。

古典的テスト理論の問題点例

□ 改訂のしにくさ

性格検査や質問紙尺度の数項目を別の新しい項目と入れ替える時、テスト全体を多数の受験者に実施して改めて評価基準を設定しなければならない。

→テストを容易に改訂することができない。

□ 古典的テスト理論の基本となる式

被験者 p のテスト得点 $x_p =$ 真値成分 $\tau_p +$ 測定誤差 e_p

→数学的に証明できない, 実験的にも確かめられない, 前提として受け入れるしかない公式。

項目反応理論(item response theory, IRT)

- テスト全体と被験者集団を対象とした古典的テスト理論と異なり、**1人の被験者がテストの1項目に応答する際の正解率**をモデル化。

- **項目特性関数** $p(\theta)$

テスト項目に被験者が正答する確率を被験者の特性値 θ の関数として表現したもの。

→項目特性関数の数学的表現の違いにより、様々な項目反応モデルが生じる。

→ちなみに、グラフ化したものを**項目特性曲線**という。

色々な項目反応モデル

- ロジスティック・モデル(logistic model)
- 正規累積モデル(normal ogive model)
- ラッシュモデル(Rasch model)
- 段階反応モデル(graded response model)
- 部分的反応モデル(partial credit model)
- 名義反応モデル(nominal response model)
- 連続的反應モデル(continuous response model)
- 拡張ロジスティックモデル(generalized logistic model)
- 一般正規累積モデル(general normal-ogive model)

ロジスティック・モデル

- 色々なモデルがある。一般に, ロジスティック・モデルが使われる。数学的な取扱いが容易。
- ロジスティック・モデルで扱うパラメータ
 1. 識別力(a)…被験者の特性値(能力)の違いが正答確率にどの程度反映するか。
 2. 困難度(b)…大きいほどテスト項目が難しい問題である。
 3. 当て推量(c)…被験者が偶然正答する確率。

ロジスティック・モデル

- 1パラメタ・ロジスティックモデル(one-parameter logistic model)
項目の特徴のうち、**困難度のみ**をモデル化したもの。困難度は項目ごとに定まるが、全ての項目の識別力パラメタは等しいと考える。
- 2パラメタ・ロジスティックモデル(two-parameter logistic model)
項目の**困難度**と**識別力**によってテスト項目の正答確率を表現しようとするモデル。
- 3パラメタ・ロジスティックモデル(three-parameter logistic model)
項目の**困難度**と**識別力**に加えて、**当て推量**により被験者が偶然正答する確率についてのパラメタ c を導入している。
- 4パラメタ・ロジスティックモデル(four-parameter logistic model)
不注意などによって、能力が極めて高い被験者もテスト項目に間違える可能性が残るため、**天井**のパラメタを加え、 θ が大きくなっても正答確率が最大で1未満となるようにした。

2パラメタ・ロジスティックモデル

正答確率

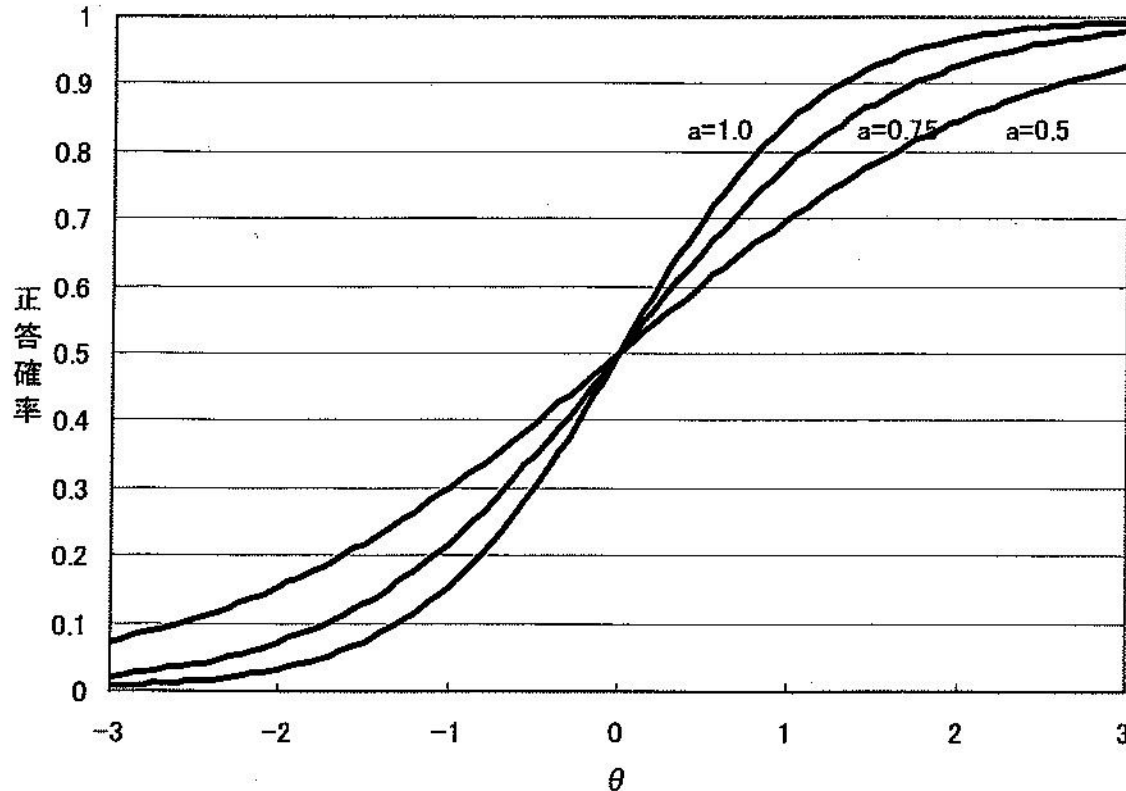


図2. 2母数ロジスティックモデルの項目特性曲線

特性値

(長崎・萩原, 2004)

項目反応理論を使ったテストの作り方

1. テストの計画
2. 項目の作成
3. 予備テスト
4. 予備テストの結果の分析
5. 項目パラメータ値の再推定
6. 項目パラメータ値の等化
7. テストの編集
8. テストの標準化
9. テストの修正

手順1 計画～予備テスト

1. テストの計画

2. 項目の作成

3. 予備テスト

4. 予備テストの結果の分析

5. 項目パラメータ値の再推定

6. 項目パラメータ値の等化

7. テストの編集

8. テストの標準化

9. テストの修正

1. 理論的な検討を十分に行い、測定内容や領域、実施方法等について計画を立てる。
2. 実際のテストで用いる項目を作成する。
3. 作成された項目が意図通りに機能するかどうかを確認するため、予備テストを行う。

手順2 集計～除去

1. テストの計画
2. 項目の作成
3. 予備テスト
4. 予備テストの結果の分析
5. 項目パラメータ値の再推定
6. 項目パラメータ値の等化
7. テストの編集
8. テストの標準化
9. テストの修正

- スクリーニング…極短に通過率の高い項目や低い項目，無答が多い被験者などを除外。
- 特性値を推定する。各項目のパラメータ値を推定する。 ※推定法:PROX推定法, 条件付き最尤推定法, 同時最尤推定法, 周辺最尤推定法, ベイズ推定法 など
- 項目を篩いにかける。
 - 識別力…推定値が極端に小さい項目(例: ≤ 0.2)
 - 困難度…推定値が極端に大きい項目(例: > 3.0)
や小さい項目(例: ≤ -3.0)
 - モデルとの適合度の悪い項目

手順2 集計～除去

1. テストの計画
2. 項目の作成
3. 予備テスト
4. 予備テストの結果の分析
5. 項目パラメータ値の再推定
6. 項目パラメータ値の等化
7. テストの編集
8. テストの標準化
9. テストの修正

□ 情報関数(information function)

古典的テスト理論で言うところの信頼性。項目反応理論では、特性値 θ のレベルごとに推定値の精度の評価をすることができる。そのような精度の指標を情報関数という。

□ 情報関数の値

情報量，テスト情報量，項目情報量などとよばれる。

□ 情報関数を曲線で表したもの

情報曲線，テスト情報曲線，項目情報曲線などとよばれる。

途中ですが、いきなり実習。

テストの項目像を見てみよう。(ソフト:R)

1. 項目特性曲線を描いてみよう。
2. テスト特性曲線を描いてみよう。

項目特性曲線とテスト特性曲線

□ 項目特性曲線(item characteristic curve, ICC)

潜在特性 θ を独立変数として、確率関数を表したものの。特性値が θ の人が各項目へ正答する確率(期待値)を表したものの。(y軸に正答確率をとる)

□ テスト特性曲線(test characteristic curve, TCC)

テスト特性関数(テスト反応関数)をグラフにしたものの。特性値が θ の人が何度もテストを受けた時の平均得点の推移を表したものの。(y軸に正答した項目数をとる)

パッケージ“Irtoys”をインストール

インストールする

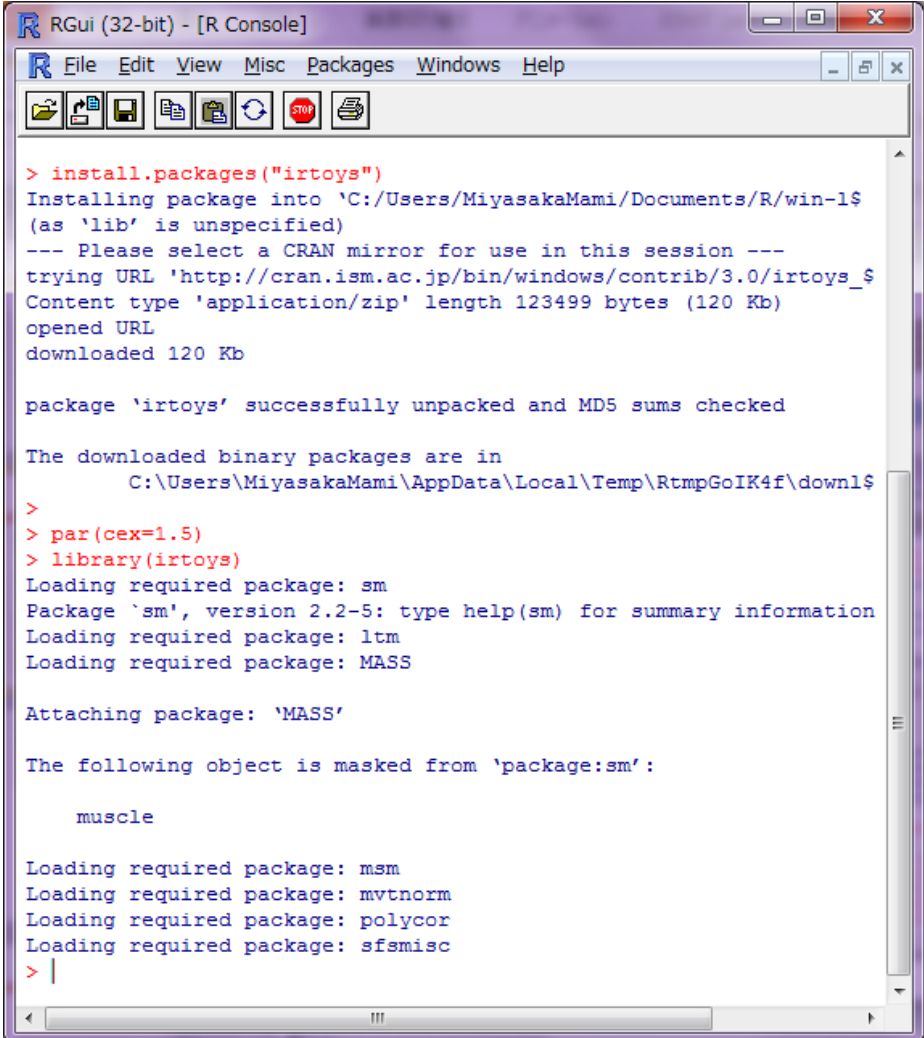
```
install.packages("irtoys")
```

文字の大きさを標準の1.5倍にしたかったらする

```
par(cex=1.5)
```

読みこむ

```
library(irtoys)
```



```
RGui (32-bit) - [R Console]
File Edit View Misc Packages Windows Help

> install.packages("irtoys")
Installing package into 'C:/Users/MiyasakaMami/Documents/R/win-l$
(as 'lib' is unspecified)
--- Please select a CRAN mirror for use in this session ---
trying URL 'http://cran.ism.ac.jp/bin/windows/contrib/3.0/irtoys_$
Content type 'application/zip' length 123499 bytes (120 Kb)
opened URL
downloaded 120 Kb

package 'irtoys' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Users\MiyasakaMami\AppData\Local\Temp\RtmpGoIK4f\downl$
>
> par(cex=1.5)
> library(irtoys)
Loading required package: sm
Package `sm`, version 2.2-5: type help(sm) for summary information
Loading required package: ltm
Loading required package: MASS

Attaching package: `MASS`

The following object is masked from `package:sm`:

  muscle

Loading required package: msm
Loading required package: mvtnorm
Loading required package: polycor
Loading required package: sfsmisc
> |
```

この項目特性曲線を描くつもり

	識別力	困難度	当て推量	備考
	a_i	b_i	c_i	
問1	0.5	-2.0	0.0	識別力も困難度も低い
問2	2.0	2.0	0.0	識別力も困難度も高い
問3	2.0	-2.0	0.0	識別力は高く, 困難度は低い
問4	0.5	2.0	0.0	識別力は低く, 困難度は高い

項目特性曲線を描く

パラメータを入力する

```
ip2<-matrix(c(
```

```
# aj bj cj
```

```
0.5, -2.0, 0.0, # 識別力も困難度も低い
```

```
2.0, 2.0, 0.0, # 識別力も困難度も高い
```

```
2.0, -2.0, 0.0, # 識別力は高く, 困難度は低い
```

```
0.5, 2.0, 0.0), ncol=3, byrow=T) #識別力は低く困難度は高い
```

正答確率を計算し, 項目特性曲線をプロットする:plot(irf)

```
plot(irf(ip2,x=seq(-6.0, 6.0, length=1000)))
```

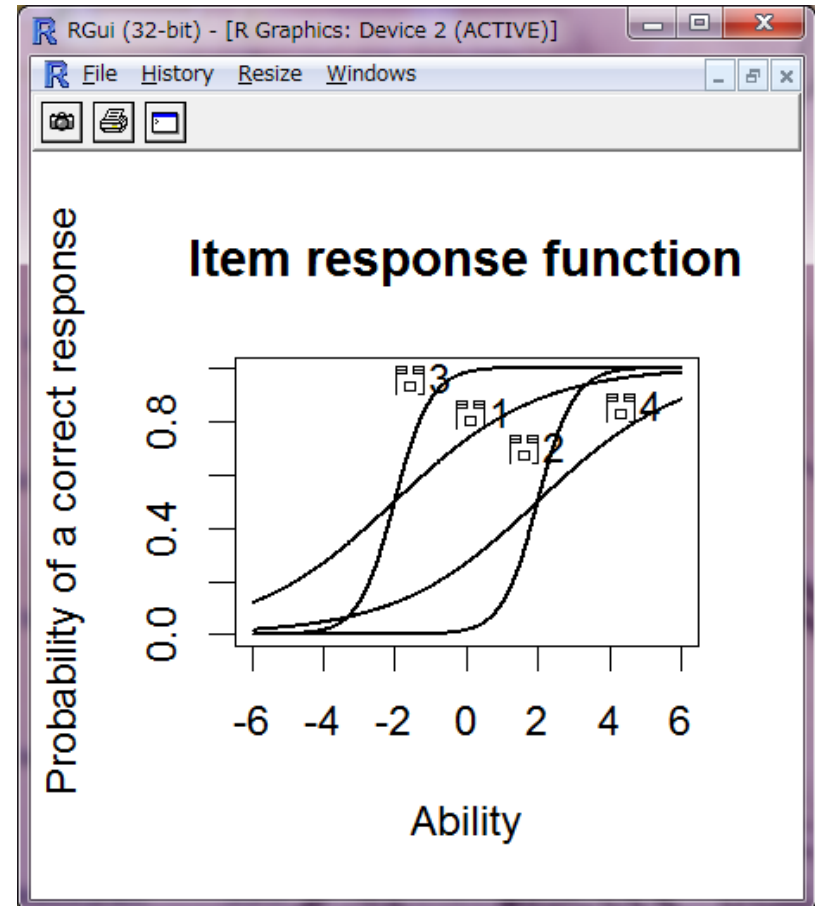
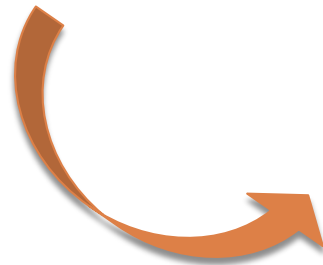
データ系列名を表示させる

```
text(locator(4),c("問1", "問2", "問3", "問4"))
```

項目母数をip2へ代入する。3パラメタ・ロジスティックモデルの場合, cjにも数字が入る。2パラメタ・ロジスティックモデルはcjを0としたモデル。

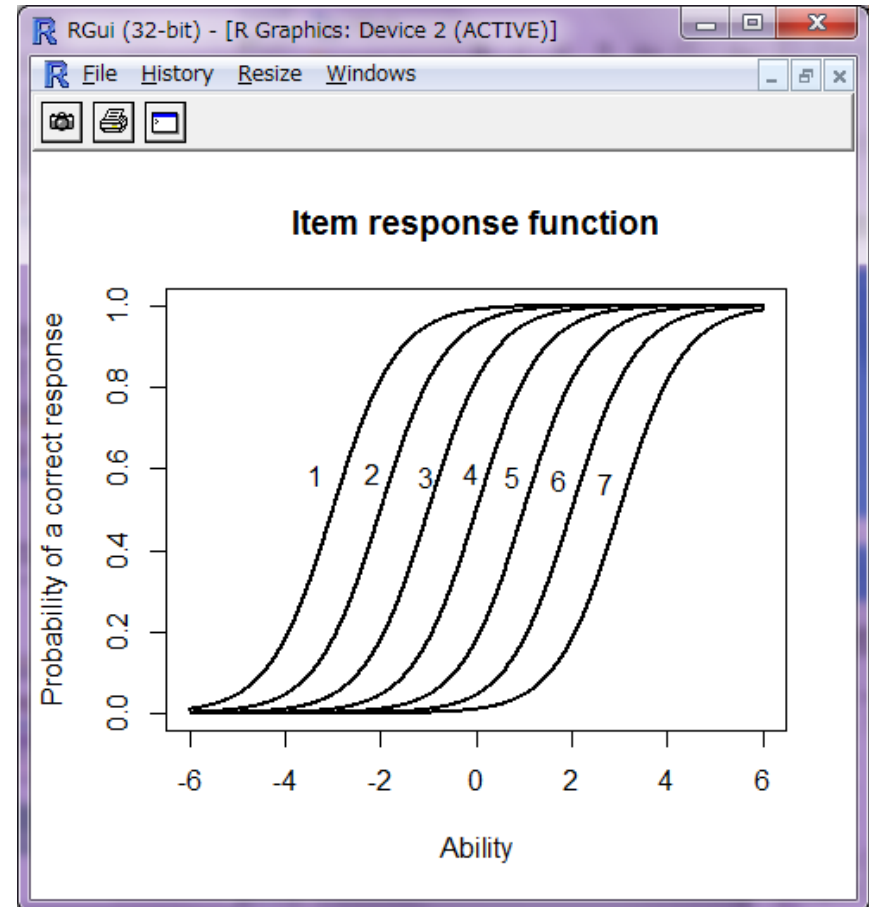
グラフが出来たら、「『問1』はここ」という具合に、
系列名を書きたいところを順にクリックする。

```
RGui (32-bit) - [R Console]
File Edit View Misc Packages Windows Help
[Icons]
> ip2<-matrix(c(
+ # aj bj cj
+ 0.5, -2.0, 0.0,
+ 2.0, 2.0, 0.0,
+ 2.0, -2.0, 0.0,
+ 0.5, 2.0, 0.0),
+ ncol=3, byrow=T)
>
> plot(irf(ip2,x=seq(-6.0, 6.0, length=1000)))
> text(locator(4),c("問1","問2","問3","問4"))
> |
```



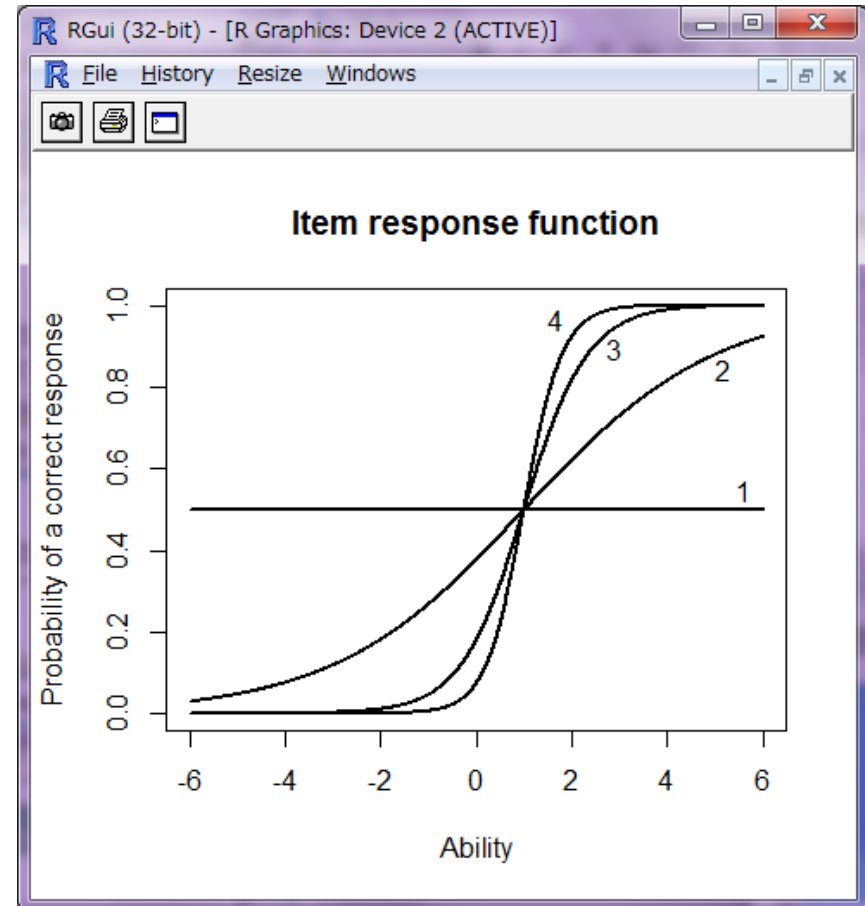
Column1: 識別力を一定に, 困難度を変えると

```
ip2<-matrix(c(
# aj bj cj
1.5, -3.0, 0.0,
1.5, -2.0, 0.0,
1.5, -1.0, 0.0,
1.5, 0.0, 0.0,
1.5, 1.0, 0.0,
1.5, 2.0, 0.0,
1.5, 3.0, 0.0), ncol=3, byrow=T)
plot(irf(ip2,x=seq(-6.0, 6.0,
+ length=1000)))
text(locator(7),c("1","2","3",
+ "4","5","6","7"))
```



Column2: 困難度を一定に，識別力を変えると

```
ip2<-matrix(c(
# aj bj cj
0.0, 1.0, 0.0,
0.5, 1.0, 0.0,
1.5, 1.0, 0.0,
2.5, 1.0, 0.0), ncol=3, byrow=T)
plot(irf(ip2,x=seq(-6.0, 6.0,
length=1000)))
text(locator(4),c("1","2","3",
+ "4"))
```

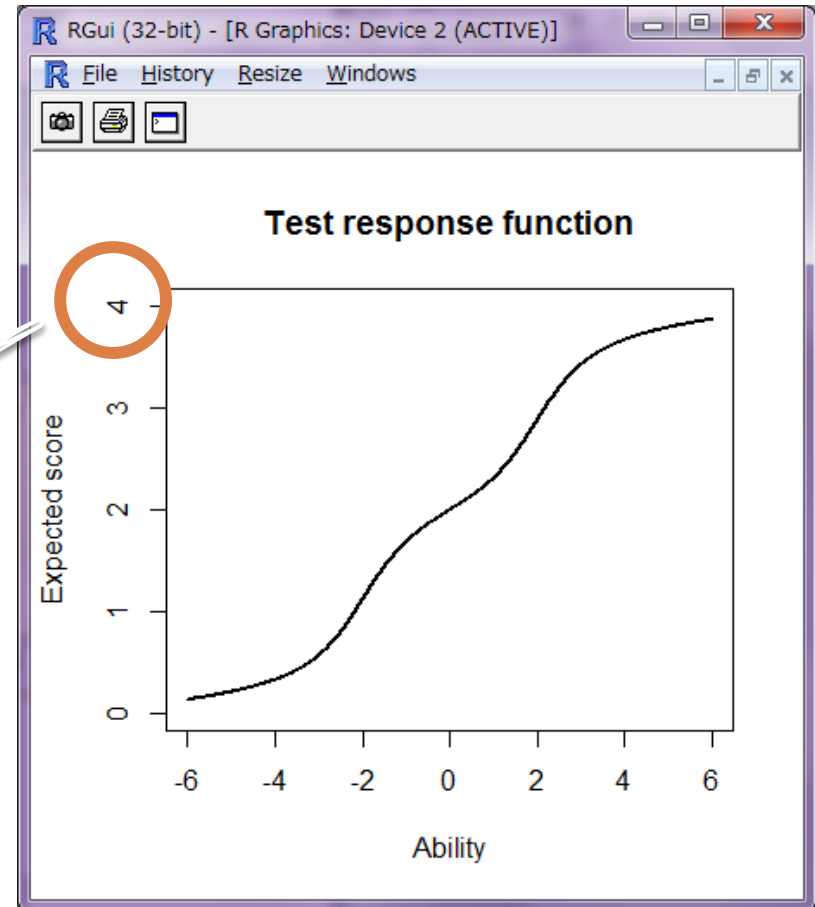


テスト特性曲線を描く

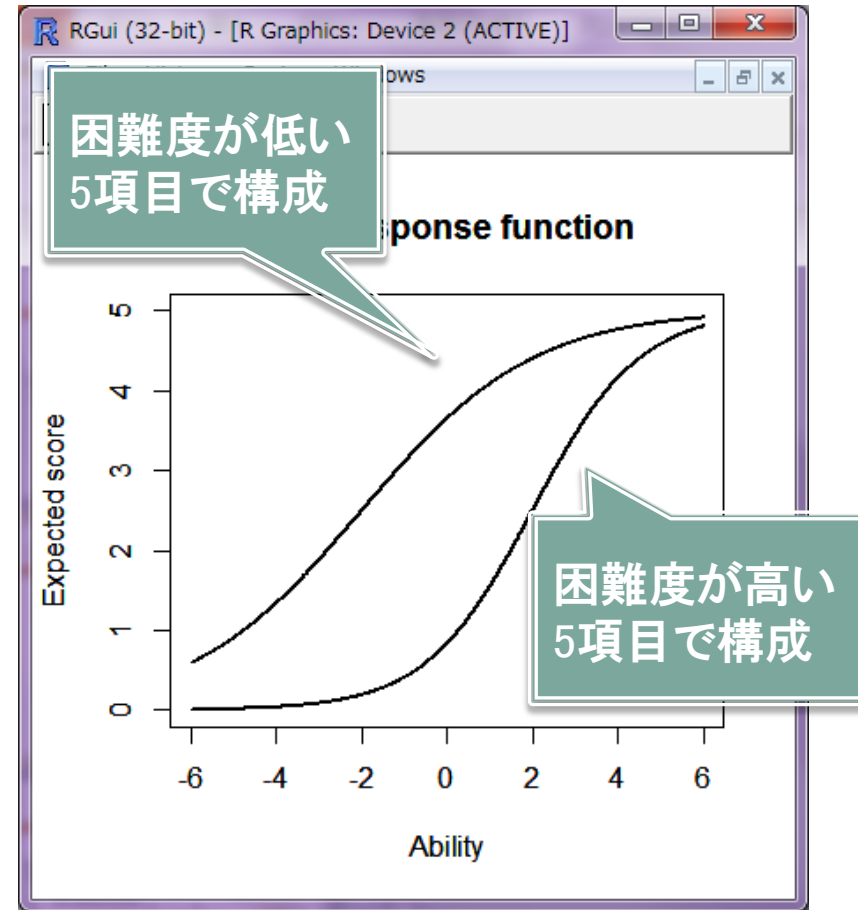
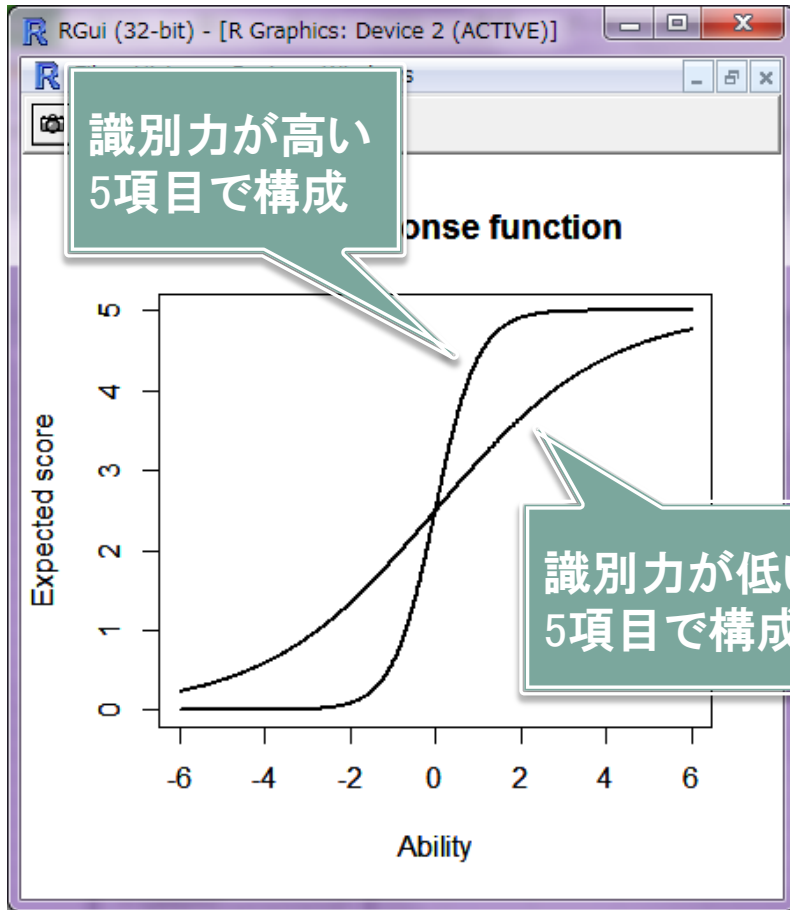
テスト特性関数を描く: `plot(trf)`

```
plot(trf(ip2,x=seq(-6.0, 6.0,  
+ length=1000)))
```

満点

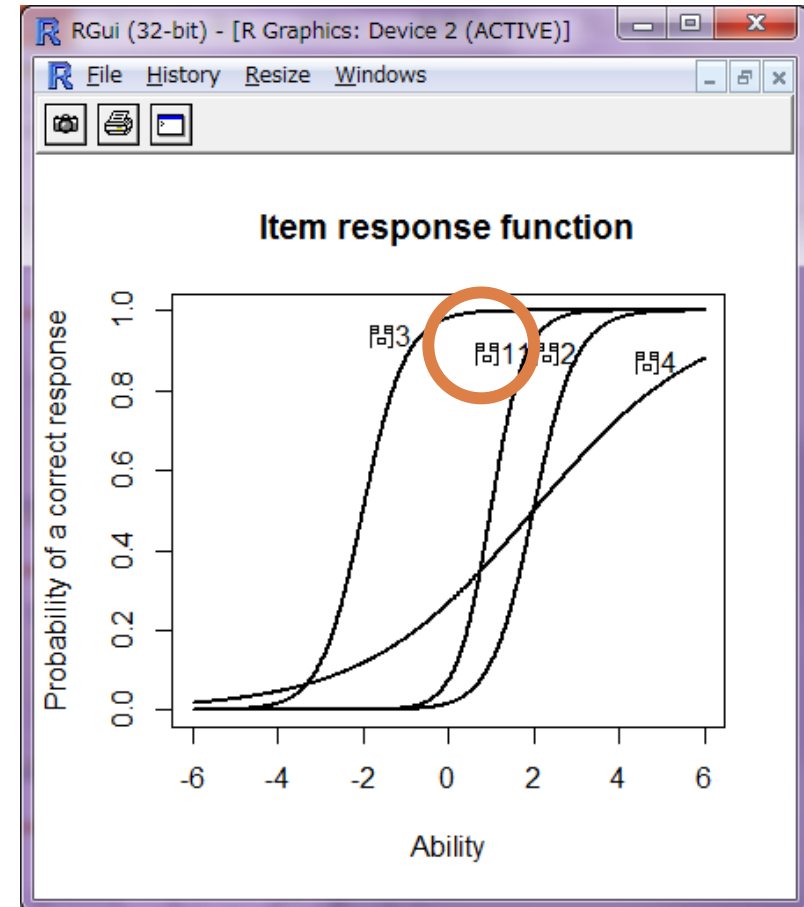
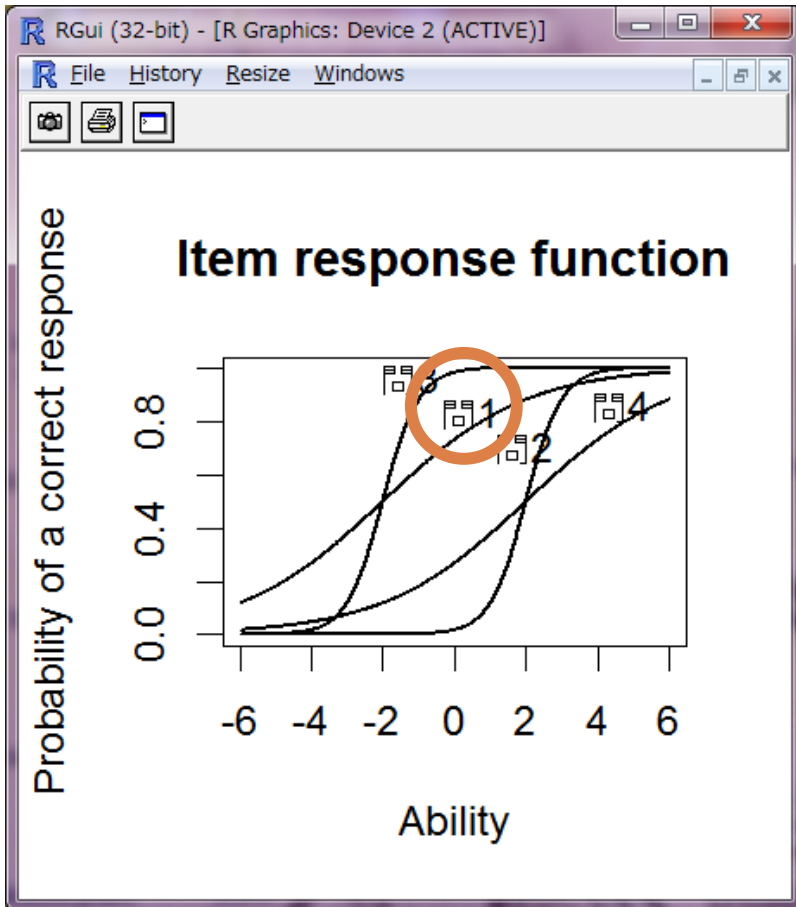


Column3: テスト特性曲線比較



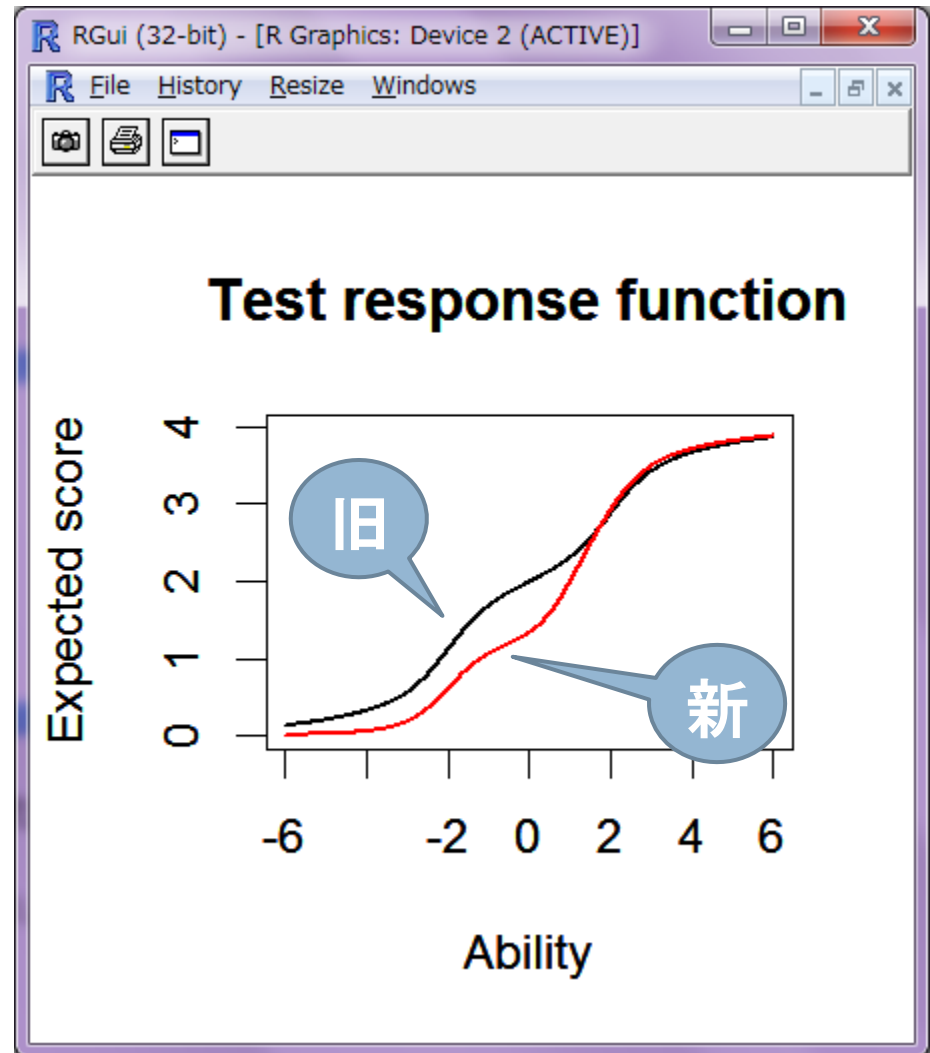
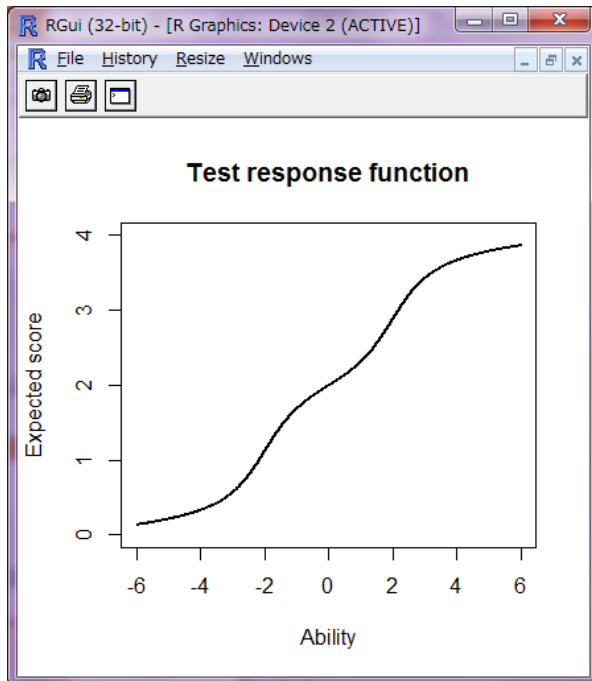
問1を変えてみる(識別力2.5,困難度1.0にしてみた)

□ 項目特性曲線は...



問1を変えてみる(識別力2.5,困難度1.0にしてみた)

□ テスト特性曲線を重ねてみると…

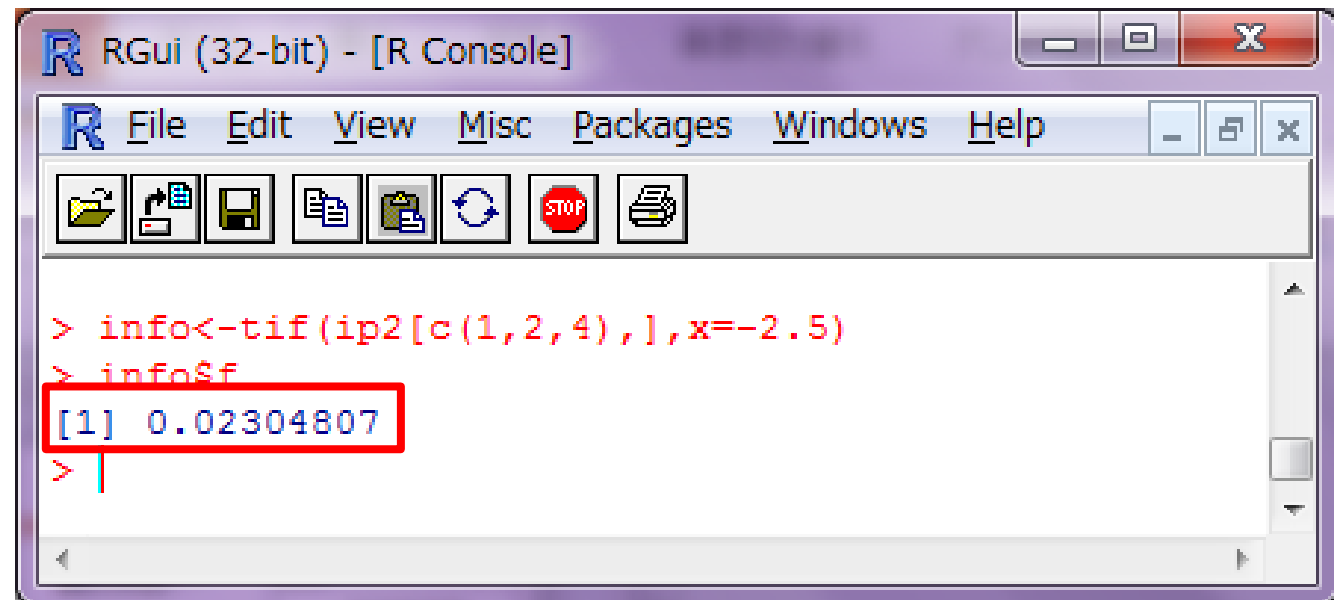


Column4: 情報量

θ が -2.5 である受験者Aが問1, 問2, 問4に解答した時の、
被験者Aのテスト情報量を出す。: `tif()`

```
info <- tif(ip2[c(1,2,4),], x = -2.5)
```

```
info$f
```

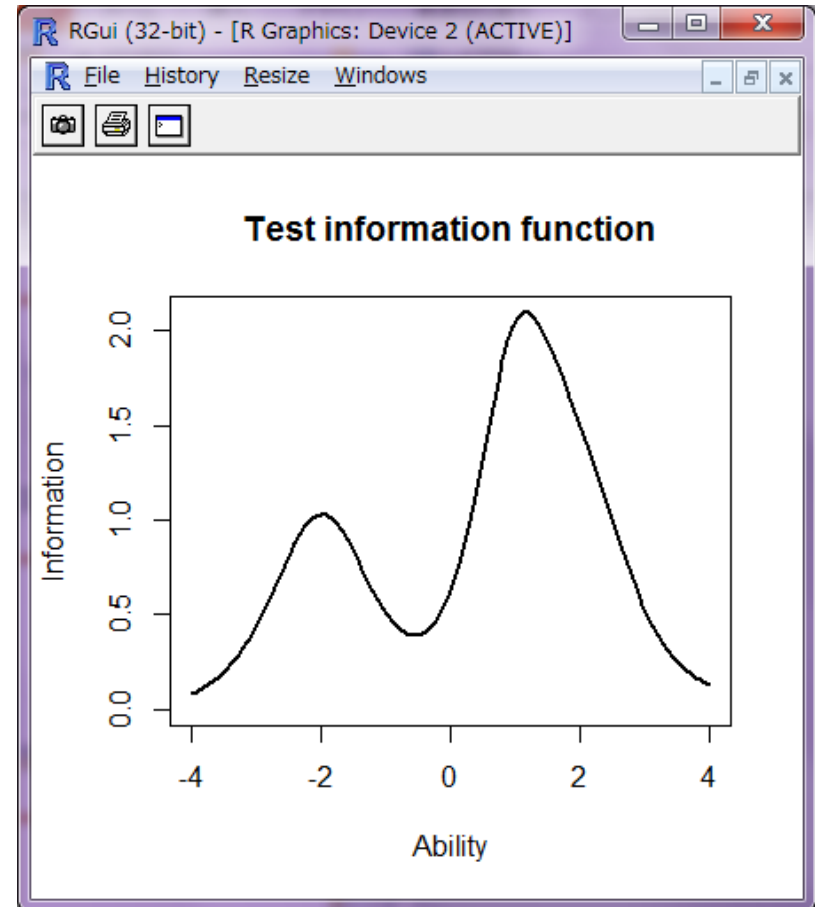


```
RGui (32-bit) - [R Console]
File Edit View Misc Packages Windows Help
[Icons: Run, Copy, Paste, Undo, Redo, Stop, Print]
> info <- tif(ip2[c(1,2,4),], x = -2.5)
> info$f
[1] 0.02304807
> |
```

Column5: 情報曲線

テスト情報関数を描画する。:plot(tif)

plot(tif(ip2))



はい、講義再開。

余計なものを除去したところから・・・

手順3 再推定

1. テストの計画
2. 項目の作成
3. 予備テスト
4. 予備テストの結果の分析
5. 項目パラメータ値の再推定
6. 項目パラメータ値の等化
7. テストの編集
8. テストの標準化
9. テストの修正

適当と判断された項目について、改めてパラメータ値を推定する。

↓ なぜなら

一部の項目を除去したことが他の項目のパラメータ推定値にも影響を及ぼすことがあるから。

手順4 等化

これは奥が深い。
45分では説明できないのでパス。

1. テストの計画
2. 項目の作成
3. 予備テスト
4. 予備テストの結果の分析
5. 項目パラメータ値の再推定
6. 項目パラメータ値の等化
7. テストの編集
8. テストの標準化
9. テストの修正

1. 複数の版が存在する。
2. 各版ごとに推定された項目パラメータ値を相互に比較したい。

↓ならば

全ての項目パラメータ値を、それが属する版に関わらず、共通の原点と単位とをもつ共通尺度上で表わさなければならない。

※複数の版に含まれる全ての項目が1つの項目プールを構成することになる

手順5 編集

1. テストの計画
2. 項目の作成
3. 予備テスト
4. 予備テストの結果の分析
5. 項目パラメータ値の再推定
6. 項目パラメータ値の等化
7. **テストの編集**
8. テストの標準化
9. テストの修正

- 測定目的に応じて、項目プール中から適切な特性をもつ項目を選んで編集される。
- 難易度水準の異なる複数の版，ほぼ平行な複数の版も作れる。

手順6 標準化

1. テストの計画
2. 項目の作成
3. 予備テスト
4. 予備テストの結果の分析
5. 項目パラメータ値の再推定
6. 項目パラメータ値の等化
7. テストの編集
8. テストの標準化
9. テストの修正

- 新しく編集したテストを更に多数の被験者に対して実施した時には、その結果をもとに各項目のパラメータ値の推定と等化を再度行う。
- 各被験者の測定結果は推定尺度値で表される。推定尺度値から何らかの有効な情報を引き出すためには数値を解釈するため、パーセンタイル順位やT得点などとの対応関係を与える**基準表**を作る。

↓つまり

標準化 = 基準表を作る一連の手続き

手順7 修正

1. テストの計画
2. 項目の作成
3. 予備テスト
4. 予備テストの結果の分析
5. 項目パラメータ値の再推定
6. 項目パラメータ値の等化
7. テストの編集
8. テストの標準化
9. テストの修正

□ 修正を加える場合

新しいテストの精度はすでに推定されている項目パラメータ値を用いてテスト情報量を計算すれば良い。新たにデータをとる必要はない。

□ 版の一部の項目を取り換えた

新しい項目が同じ項目プールに含まれていた場合、標準化のデータを集める必要はない。従来の基準表を用いて推定尺度値の解釈を行う。

Reference

- 芝祐順(編):項目反応理論—基礎と応用—. 東京大学出版会, 東京, 1991.
- 長崎栄三・萩原康仁:算数達成度の項目反応理論による比較分析. 国立教育政策研究所, 東京, 2004.
- 服部環:心理・教育のためのRによるデータ解析. 福村出版, 東京, 2011.
- 村木英治:シリーズ<行動計量の科学>項目反応理論. 朝倉書店, 東京, 2011.