



Rによる心理学研究法入門

5章 教育測定に関する実証研究

2015/07/15(水)心理データ解析演習
D1柘田恵・ D1宮坂まみ

Overview

- 古典的テスト理論の解説 [宮坂]
- 安永ら(2012)の解説 [柘田]
- Rを用いた分析の実習
 - 合計得点の算出 [柘田]
 - 合計得点に基づいた群分け [柘田]
 - 古典的テスト理論における項目分析を用いた値の算出
 - 項目難易度 (項目得点率) の算出 [柘田]
 - 解答類型分類率の算出 [柘田]
 - 項目識別力(I-T相関)の算出 [宮坂]
 - 得点率の差およびI-T相関の差に関する統計的推定 [宮坂]
 - 補足 [宮坂]
 - まとめ [宮坂]

Overview

- **古典的テスト理論の解説 [宮坂]**
- 安永ら(2012)の解説 [杓田]
- Rを用いた分析の実習
 - 合計得点の算出 [杓田]
 - 合計得点に基づいた群分け [杓田]
 - 古典的テスト理論における項目分析を用いた値の算出
 - 項目難易度 (項目得点率) の算出 [杓田]
 - 解答類型分類率の算出 [杓田]
 - 項目識別力(I-T相関)の算出 [宮坂]
 - 得点率の差およびI-T相関の差に関する統計的推定 [宮坂]
 - 補足 [宮坂]
 - まとめ [宮坂]

古典的テスト理論

- テスト：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
- 古典的テスト理論 (Classical Test Theory: CTT)
 - 受検者 p のテスト得点 $x = \text{真値成分}\tau_p + \text{誤差}e_p$ を想定
 - 信頼性の検討，妥当性の検討，項目分析などを行う

古典的テスト理論

- 項目分析 (item analysis)
 - テストを構成している個々の項目が期待している機能を果たしているかどうか
 - 古典的テスト理論に基づく項目分析
 - **項目困難度**
ある項目を解いた受検者のうち何人が正解したか
(= 通過率, 平均項目得点)
 - **識別力** (弁別力)
合計得点の高い受検者と低い受検者を弁別できるか

古典的テスト理論

補足：

古典的テスト理論は、いくつか問題点が指摘されている

正答数（素点）がテストそのものに依存する

結果が被験者集団の分布に依存する

基本となる式（テスト得点 $x = \text{真値}\tau_p + \text{誤差}e_p$ ）に根拠がない

→その点を解決しているのが**項目反応理論(現代テスト理論)**

興味のある方は楠見先生・高橋先生の2013年度心理データ解析演習、「項目反応理論」（担当：宮坂）をご参照ください。前半で古典的テスト理論との関連をまとめてあります。

それでも古典的テスト理論は今も使われている。

状況に応じて使い分けると良い。今回は古典的テスト理論。

古典的テスト理論と項目反応理論

- 項目反応理論と比べたときの古典的テスト理論の利点
 - ①簡単な公式でテストの性能評価ができるため、応用範囲が広い（例えば、IRTのための事前分析にも有効）
 - ②テストや項目の評価に用いる受検者集団が、そのテストが対象としている母集団をよく代表している場合には十分機能する。
→対象母集団がはっきりしていて、受検者層が想定しやすく、ぶれない場合は有効に機能
 - ③テストや項目の性能評価のためのサンプル数が少なくて済む

Overview

- 古典的テスト理論の解説 [宮坂]
- **安永ら(2012)の解説 [柘田]**
- Rを用いた分析の実習
 - 合計得点の算出 [柘田]
 - 合計得点に基づいた群分け [柘田]
 - 古典的テスト理論における項目分析を用いた値の算出
 - 項目難易度 (項目得点率) の算出 [柘田]
 - 解答類型分類率の算出 [柘田]
 - 項目識別力(I-T相関)の算出 [宮坂]
 - 得点率の差およびI-T相関の差に関する統計的推定 [宮坂]
 - 補足 [宮坂]
 - まとめ [宮坂]

研究の概要

- テストは日本の学校教育で重要！
 - テスト作成に関する実証的知見の不足；専門家の知識・経験依存
- 受検者に対して適切な評価はできている？
- テストを構成する性質
 - ①内容的性質（内容的妥当性）：「何」を問うか
 - 教科，範囲，項目内容
 - ②**構造的性質**：「どのように」問うか
 - テストフォーマット，設問形式，設問の問い方
- * 同じ内容の設問でも，設問の設定の仕方により受検者の回答は変化する

研究の概要

＜本研究の目的＞

- 設問の構造的性質を評価するために、中学生を対象に国語のテストを実施
 - 読解プロセス
 - 回答欄の字数制限
 - 空所の表記法
 - 一文抜き出し問題

研究の概要

<具体的な手続き>

- 参加者：愛知県(2校)と三重県の公立中学校に所属する中学3年生493名(男子252名, 女子241名)
- 回答時間：50分
- 問題本文：「和の思想」長谷川權
(内容) 日本の間についての西洋と対比した説明文
(設問) 記述式8問, 多肢選択式7問の計15問

研究の概要

＜本研究で研究対象とした設問の概要＞

設問	概要	形式
1b	傍線部分⑤の「心理的な間」に関して具体例を挙げて説明	記述
5	傍線部分④の日本の家の特徴について西洋の家と比較し、本文の言葉を使って記述	記述
7a	本文の内容に関する会話文中の空所アに当てはまる語句を記述する	記述
8	傍線部分⑥の「和が成り立つ」と言える理由を本文中から一文抜き出す	記述

1b: 読解のプロセス
5: 回答欄の字数制限
7a: 空所の表記法
8: 一文抜き出し問題

- 4つの設問は要因が交絡しないように組み合わせ、8種類の問題冊子を作成(個々の生徒にランダムに割り当て)
- 問題冊子間で受検者の国語能力にも差なし

＜本研究で取り上げる設問の操作＞

設問1b: 正答となる具体例の有無を操作

A条件: 同じ段落内に設問の答えとなる具体例がない本文(統合・解釈)

B条件: 同じ段落内に設問の答えとなる具体例がある本文(情報へのアクセス・取り出し)

設問5: 回答欄の文字数を操作

A条件: 45字以上50字以内

B条件: 55字以内

C条件: 字数制限を行わない

研究の概要

設問の解答類型

- 評定は第一筆者が解答類型に従って行い，判断の迷う箇所は第二筆者と合議の上，評定

*解答類型：回答に対してあらかじめ定められた評価基準

*本研究では，類型の数字にrをつけたものを項目得点(r_1, r_2, \dots, r_0)

設問	内容	類型	得点
1b	自分で具体例を挙げている	1 正答	1
	本文中から具体例を挙げている(B条件のみ)	2 正答	1
	本文中から具体性に欠ける文章を選んでいる	3 準正答	0.5
	上記以外の回答	9 誤答	0
	回答の記述なし	0 無回答	0

設問	内容	類型	得点
5	正答に求められる内容 ①②両方の内容に着目した文を書いている	1 正答	1
	①の内容のみに着目した文を書いている	2 準正答	0.5
	②の内容のみに着目した文を書いている	3 準正答	0.5
	上記以外の回答	9 誤答	0
	回答の記述なし	0 無回答	0

Overview

- 古典的テスト理論の解説 [宮坂]
- 安永ら(2012)の解説 [柘田]
- Rを用いた分析の実習
 - **合計得点の算出 [柘田]**
 - 合計得点に基づいた群分け [柘田]
 - 古典的テスト理論における項目分析を用いた値の算出
 - 項目難易度 (項目得点率) の算出 [柘田]
 - 解答類型分類率の算出 [柘田]
 - 項目識別力(I-T相関)の算出 [宮坂]
 - 得点率の差およびI-T相関の差に関する統計的推定 [宮坂]
 - 補足 [宮坂]
 - まとめ [宮坂]

合計得点の算出

- > #データファイルの読み込み
- > saiten <- read.csv("saiten.csv")
- > #最初の6行を表示
- > head(saiten)

```
> head(saiten)
  id e1 e5 d1a d1b d2a d2b d3 d4 d5 d6a d6b d6c d7a d7b d7c d8 d9 s1a s1b s2a s2b s3 s4 s5 s6a s6b s6c s7a
1  1  B  C  r2  r9  r1  r5  r2  r1  r2  r3  r1  r4  r9  r2  r2  r1  r2  1.0  0.0  0  0  1  1  0.5  0  1  1  0
2  2  B  C  r2  r2  r2  r5  r2  r1  r1  r7  r1  r4  r1  r9  r1  r3  r1  1.0  1.0  1  0  1  1  1.0  0  1  1  1
3  3  B  C  r3  r9  r4  r3  r2  r9  r0  r6  r4  r3  r9  r1  r1  r9  r4  0.5  0.0  0  0  1  0  0.0  1  0  0  0
4  4  B  C  r2  r3  r2  r5  r2  r1  r1  r6  r7  r4  r0  r2  r0  r0  r3  1.0  0.5  1  0  1  1  1.0  1  0  1  0
5  5  B  C  r3  r2  r2  r5  r2  r9  r2  r6  r1  r4  r9  r9  r9  r9  r1  0.5  1.0  1  0  1  0  0.5  1  1  1  0
6  6  B  C  r2  r3  r2  r5  r2  r1  r1  r6  r1  r7  r1  r1  r2  r1  r4  1.0  0.5  1  0  1  1  1.0  1  1  0  1
  s7b s7c s8 s9
1  0  1  1  0
2  0  1  1  1
3  1  1  0  0
4  0  0  0  0
5  0  0  0  1
6  1  1  1  0
```

e1:設問1bの条件 (A, B)

e5:設問5の条件 (A, B, C)

d1a – d9: 設問1aから設問9までの解答類型による評定値 (r0:類型0, ..., r9: 類型9)

s1a – s9:設問1aから設問9までの項目得点

合計得点の算出と パーセンタイル値の算出

>#合計得点の算出 (検討対象のs1b, s5, s7a,s8は条件で異なるため, 除外)

```
> saiten$goukei <- with(saiten, s1a+s2a+s2b+s3+s4+s6a+s6b+s6c+s7b+s7c+s9)
```

- 古典的テスト理論に基づいて
→低群27%, 中群46%, 高群27%
#パーセンタイルの値の算出



```
> quantile(saiten$goukei, prob = c(0.27, 0.73))
```

* c()内に指定するパーセンタイル値

27% 73%

4.0 7.5



低群	4.5点未満(4.0点以下)
中群	4.5点以上7.5点未満
高群	7.5点以上

quantileのデフォルトは, 25%刻み
- 0%点, 25%点,
50%点, 75%点,
100%点

Overview

- 古典的テスト理論の解説 [宮坂]
- 安永ら(2012)の解説 [柘田]
- Rを用いた分析の実習
 - 合計得点の算出 [柘田]
 - **合計得点に基づいた群分け [柘田]**
 - 古典的テスト理論における項目分析を用いた値の算出
 - 項目難易度 (項目得点率) の算出 [柘田]
 - 解答類型分類率の算出 [柘田]
 - 項目識別力(I-T相関)の算出 [宮坂]
 - 得点率の差およびI-T相関の差に関する統計的推定 [宮坂]
 - 補足 [宮坂]
 - まとめ [宮坂]

合計得点に基づいた群分け

cut () : 群分けに使用する関数

`cut (変数名, right = , breaks=c(分割点), label=c(カテゴリ名 (群名)), ordered_result=TRUE)`

- `right=FALSE` : 分割点の右端を含まない
←`right=TRUE`: 分割点の右端を含む
- 分割点は, `-Inf, ..., Inf` (`-Inf`: 最小値, `Inf`: 最大値) で指定
- 群の数 : `-Inf, Inf` を含んだ分割点 - 1
→ 分割点で分割される区間分のカテゴリ名を指定
- `ordered_result=TRUE`: 群分け変換後のデータを順序尺度に指定

合計得点に基づいた群分け

- 先ほど算出したパーセンタイル値に基づき群分け

- 27%: 4.0, 73%: 7.5



低群(L)	4.5点未満(4.0点以下)
中群(M)	4.5点以上7.5点未満
高群(H)	7.5点以上

```
cut(saiten$goukei, right=FALSE, breaks=c(-Inf, 4.5, 7.5, Inf), labels=c("L", "M", "H"), ordered_result=TRUE)
```

* right = FALSEにするために、4.0ではなく、4.5を採用

```
> #合計得点に基づいた3群分け
```

```
> saiten$gunwake <- cut(saiten$goukei, right=FALSE, breaks=c(-Inf, 4.5, 7.5, Inf), labels=c("L", "M", "H"), ordered_result=TRUE)
```

```
> #gunwake変数を表示
```

```
> head(saiten$gunwake, 20)
```

```
[1] M H M M M H H H H M M H H H M M H L M]
```

```
Levels: L < M < H
```

Overview

- 古典的テスト理論の解説 [宮坂]
- 安永ら(2012)の解説 [柘田]
- Rを用いた分析の実習
 - 合計得点の算出 [柘田]
 - 合計得点に基づいた群分け [柘田]
 - 古典的テスト理論における項目分析を用いた値の算出
 - 項目難易度 (項目得点率) の算出 [柘田]**
 - 解答類型分類率の算出 [柘田]
 - 項目識別力(I-T相関)の算出 [宮坂]
 - 得点率の差およびI-T相関の差に関する統計的推定 [宮坂]
 - 補足 [宮坂]
 - まとめ [宮坂]

項目難易度(困難度)

古典的テスト理論では、項目kの困難度 B_k は項目得点の平均 ($B_k = x_k/N$)

- B_k が小さいほど、困難度は高い
→項目の難易度は**正答率** (通過率)
- 0 ~ 1の値をとる
 - 全員が正答したやさしい項目→ 1
 - 誰も解けない難しい項目→ 0

本研究での算出方法

- 解答類型に従い、各受検者の得点を算出
正答 1点；準正答 0.5点；誤答・無回答 0点
→平均値を算出

項目難易度(困難度)

- 古典的テスト理論では、項目難易度は正答率
 - ★正答率は受検者集団が異なれば変動
 - 同じ項目を用いても集団間で項目難易度は異なり、項目難易度の値はテストを受ける集団に依存
 - ⇒項目難易度の標本依存性（識別力でも同様）
- テスト得点はテストに含まれる項目の難易度に依存
 - 項目難易度（正答率）の高い項目で構成されていたらテスト得点は高い
 - ⇒テスト得点の項目依存性

項目難易度

(補足) 項目反応理論での項目難易度

- 概念的に学力や性格などの特性の強さに上限・下限はない
→測定する項目の指標に上限下限があるのは不都合
- ★項目の難易度と特性値（学力や性格などの心理的な特性の高さや強さ）を同一の尺度へ乗せて定義
→特性値よりも困難度が小さい（やさしい項目）
 - 正答する可能性が高い
- 特性値より困難度が大きい（難しい項目）
 - 誤答する可能性が高い

* 困難度と項目反応のパターンを照合
→特性値を共通の尺度上で推定
⇒個人間で比較可

項目番号	1	2	3	4	5	6	7	8	9	10	11	特性値
困難度	-5	-4	-3	-2	-1	0	1	2	3	4	5	
個人番号	1	1	1	0	0		0					-2.5
	2	1	0	1	0			0		0		-3.5
	3				1		1	1	0	0	1	3.5
	...											

* 本来は数理モデル(ロジスティックモデル等)で特性値を推定

条件ごとの得点率(難易度)

`table()`: 度数分布の算出

- 条件ごとの人数を算出

`tapply()`: データをグループごとにまとめて処理

- 条件ごとの項目得点率/標準偏差の算出

→ `tapply(saiten$s1b, saiten$e1, mean)`

変数名

条件の変数名

関数*

*標準偏差の場合 = SD

```
> table(saiten$e1)

  A   B
247 246
> #条件ごとの項目得点率 (設問1b)
> tapply(saiten$s1b, saiten$e1, mean)
      A           B
0.1740891 0.5670732 |
> #条件ごとの標準偏差 (設問1b)
> tapply(saiten$s1b, saiten$e1, sd)
      A           B
0.2629891 0.4386589
```

```
> x<-list(age=18:28, height=150:183, weight=40:87)
> lapply(x,mean)
$age
[1] 23

$height
[1] 166.5

$weight
[1] 63.5

> sapply(x,mean)
  age height weight
23.0 166.5  63.5
```

- `apply`: 行列に一括で関数を適用
- `lapply, sapply`: 一次元のリスト型のデータの各要素に関数を適用
 - `lapply`: 結果がリスト表示
 - `sapply`: 結果を行列表示

*設問5においても、同様に求める

項目難易度(項目得点率)の算出

`subset()`: 条件ごとに必要な変数を抽出

- `subset(saiten, e1=="A", c(d1b, s1b, goukei, gunwake))`

* 採点というデータフレームの中で、`e1`がAと等しい行だけ取り出し、その中から`c()`の変数を抽出

```
> #条件ごとに必要な変数の抽出 (設問1bのA条件)
> A1 <- subset(saiten, e1=="A", c(d1b, s1b, goukei, gunwake))
> #最初の6行表示 (設問1bのA条件)
> head(A1)
  d1b s1b goukei gunwake
120  r3 0.5   6.0      M
121  r9 0.0   6.0      M
122  r3 0.5   8.0      H
123  r3 0.5   7.0      M
124  r9 0.0   4.5      M
125  r3 0.5   7.0      M
```

* 設問1のB条件, 設問5の各条件も同様に

項目難易度(項目得点率)の算出

- 各条件における群ごとの人数, 得点率, 標準偏差を算出
- table, tapplyの使用

```
> #設問5におけるA条件の群ごとの度数分布
> table(A5$gunwake)

 L  M  H
47 99 42
> #設問5におけるA条件の群ごとの得点率
> tapply(A5$s5, A5$gunwake, mean)
      L      M      H
0.2127660 0.3787879 0.5952381
> #設問5におけるA条件の群ごとの標準偏差
> tapply(A5$s5, A5$gunwake, sd)
      L      M      H
0.3417455 0.3280346 0.3533480
```

```
table(A1$gunwake) #設問1bにおけるA条件の群ごとの度数分布
table(B1$gunwake) #設問1bにおけるB条件の群ごとの度数分布
table(B5$gunwake) #設問5におけるB条件の群ごとの度数分布
table(C5$gunwake) #設問5におけるC条件の群ごとの度数分布
```

```
tapply(A1$s1b, A1$gunwake, mean) #設問1bにおけるA条件の群ごとの得点率
tapply(B1$s1b, B1$gunwake, mean) #設問1bにおけるB条件の群ごとの得点率
tapply(B5$s5, B5$gunwake, mean) #設問5におけるB条件の群ごとの得点率
tapply(C5$s5, C5$gunwake, mean) #設問5におけるC条件の群ごとの得点率
```

```
tapply(A1$s1b, A1$gunwake, sd) #設問1bにおけるA条件の群ごとの標準偏差
tapply(B1$s1b, B1$gunwake, sd) #設問1bにおけるB条件の群ごとの標準偏差
tapply(B5$s5, B5$gunwake, sd) #設問5におけるB条件の群ごとの標準偏差
tapply(C5$s5, C5$gunwake, sd) #設問5におけるC条件の群ごとの標準偏差
```

*引数を各条件のものに変え,
設問1bと設問5の残りの結果
についても算出
でも、何回も同じことするの
ちよつとめんどくさい。
→そんなときは...

項目難易度(項目得点率)の算出

★list変数の利用

```
tapply(saiten$s5, list(saiten$e5, saiten$gunwake), mean)
```

- list()の引数に条件変数と群分け変数入力
→結果を一括出力

<設問1b>

```
> tapply(saiten$s1b, list(saiten$e1, saiten$gunwake), mean)
      L      M      H
A 0.1400000 0.2037037 0.1640625
B 0.3474576 0.5504587 0.7564103
```

<設問5>

```
> tapply(saiten$s5, list(saiten$e5, saiten$gunwake), mean)
      L      M      H
A 0.2127660 0.3787879 0.5952381
B 0.3717949 0.3936170 0.6944444
C 0.2395833 0.4859155 0.7578125
```

条件ごとの群別得点率及び全体の得点率

<設問1b>

- B条件では群ごとの得点率は高群になるにつれて増加
 ⇔A条件では中群>高群 → 識別力の観点からは望ましくない
- 各群の得点率の比較：B条件>A条件

<設問5>

- A条件, C条件：低群<中群<高群
- B条件：低群≒中群<高群
- 中群・高群の得点率
 C条件で最も高い
- 低群の得点率
 B条件で最も高い

設問	条件	各群の得点率			得点率
		低群	中群	高群	
1b	A (N = 247)	.140 (.240) N = 75	.204 (.290) N = 108	.164 (.237) N = 64	.174 (.263)
	B (N = 246)	.348 (.418) N = 59	.551 (.446) N = 109	.756 (.358) N = 78	.567 (.439)
5	A (N = 188)	.213 (.342) N = 47	.379 (.328) N = 99	.595 (.353) N = 42	.386 (.360)
	B (N = 122)	.372 (.376) N = 39	.394 (.389) N = 47	.694 (.344) N = 36	.475 (.395)
	C (N = 183)	.240 (.376) N = 48	.486 (.405) N = 71	.758 (.345) N = 64	.516 (.419)

Overview

- 古典的テスト理論の解説 [宮坂]
- 安永ら(2012)の解説 [柘田]
- Rを用いた分析の実習
 - 合計得点の算出 [柘田]
 - 合計得点に基づいた群分け [柘田]
 - 古典的テスト理論における項目分析を用いた値の算出
 - 項目難易度 (項目得点率) の算出 [柘田]
 - 解答類型分類率の算出 [柘田]**
 - 項目識別力(I-T相関)の算出 [宮坂]
 - 得点率の差およびI-T相関の差に関する統計的推定 [宮坂]
 - 補足 [宮坂]
 - まとめ [宮坂]

解答類型分類率の算出

- 解答類型分類率とは，受検者の回答を評定基準である解答類型に従って振り分けた割合を示す指標

→各類型の回答を記述した受検者の割合を把握することが可能

* 解答類型分類率の算出

- `prop.table()`: 分割表の各セルの比率を算出

```
> #解答類型により分類された度数分布表 (設問1bのA条件)
> tA1 <- table(A1$d1b)
> #度数分布表を割合 (解答類型分類率) に変換 (設問1bのA条件)
> prop.table(tA1)
```

```
      r0      r1      r2      r3      r9
0.1052632 0.0242915 0.0000000 0.2995951 0.5708502
```

解答類型分類率の算出

- 同様に設問1bのB条件，設問5のA, B, C各条件について解答類型分類率を算出

```
tB1 <- table(B1$d1b) #設問1bのB条件
tA5 <- table(A5$d5) #設問5のA条件
tB5 <- table(B5$d5) #設問5のB条件
tC5 <- table(C5$d5) #設問5のC条件

prop.table(tB1) #設問1bのB条件
prop.table(tA5) #設問5のA条件
prop.table(tB5) #設問5のB条件
prop.table(tC5) #設問5のC条件
```

<設問1b>

A条件(本文中に正答の具体例なし)

→類型2で0, 類型9(誤答), 3(準正答)が多い

B条件(本文中に正答の具体例あり)

→類型2(正答)が多い, 類型9,3はAより少ない

<設問5>

C条件(字数制限なし)で類型1(正答)が最も多い

A,B条件では類型2(準正答)>類型1

この結果を表にまとめると...

表. 設問1bと設問5における条件ごとの解答類型分類率

設問	条件	解答類型分類率				
		0	1	2	3	9
1b	A	.105	.024	.000	.300	.571
	B	.114	.020	.439	.216	.211
5	A	.213	.170	.367	.064	.186
	B	.230	.287	.328	.049	.107
	C	.235	.366	.224	.077	.096

解答類型分類率の算出

- より詳細な検討：条件ごとの各群の解答分類率
 - > #設問5のA条件の群と類型のクロス集計表
 - > tcA5 <- table(A5\$d5, A5\$gunwake)
 - > #割合のクロス集計表
 - > prop.table(tcA5, 2)
 - 2:列和に対する割合, 1:行和に対する割合

```
> prop.table(tcA5, 2)
```

```
      L      M      H
r0 0.34042553 0.23232323 0.02380952
r1 0.10638298 0.12121212 0.35714286
r2 0.17021277 0.44444444 0.40476190
r3 0.04255319 0.07070707 0.07142857
r9 0.34042553 0.13131313 0.14285714
```

* 1:行和に対する割合

```
> prop.table(tcA5, 1)
```

```
      L      M      H
r0 0.40000000 0.57500000 0.02500000
r1 0.15625000 0.37500000 0.46875000
r2 0.11594200 0.63768120 0.24637680
r3 0.16666667 0.58333333 0.25000000
r9 0.45714290 0.37142860 0.17142860
```

解答類型分類率の算出

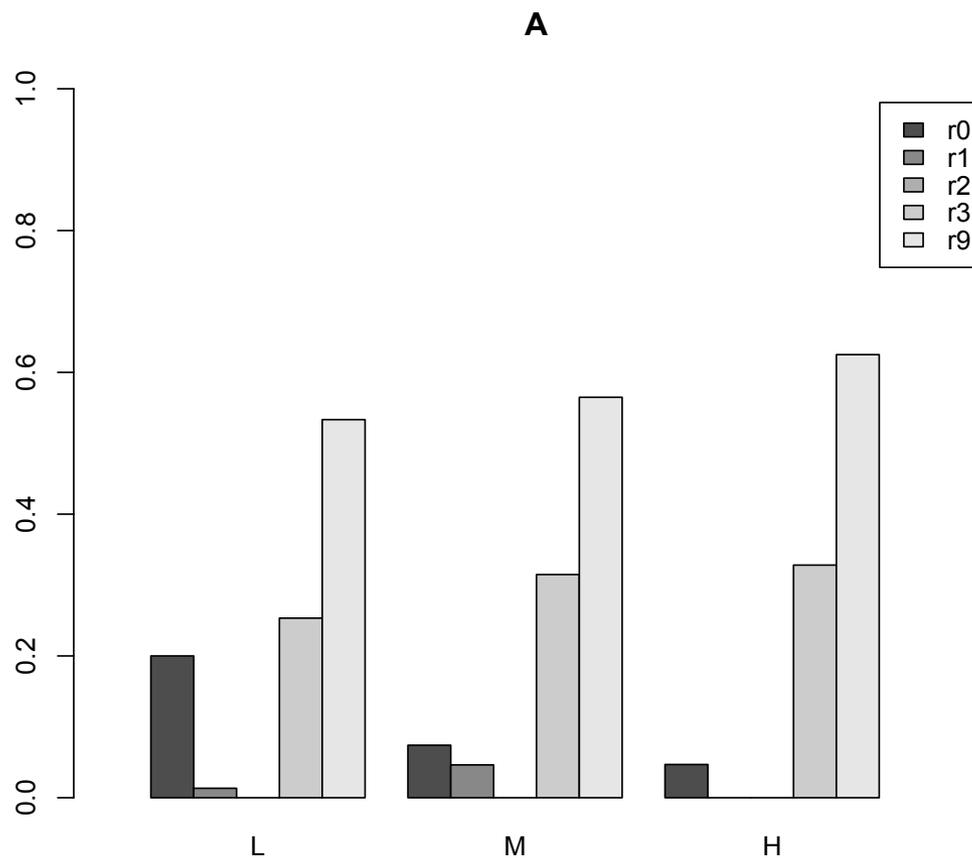
- 条件ごとの各群の解答分類率を視覚的に把握できるように棒グラフ作成

```
barplot(行列名, main=" ", xlim=c(), ylim=c(), beside=, legend=)
```

- main:図の上部中央に指定したタイトルを表示
 - * sub: 図の下部中央
- xlab, ylab: X軸, Y軸にラベル
- xlim, ylim: X軸, Y軸の表示範囲を指定
- beside:行列データに対する棒グラフの表示形式
TRUE=並列表示, FALSE=積み上げ表示
- legend = TRUE:凡例の表示

```
barplot(p.tcA5, main="A", xlim=c(0,20), ylim=c(0,1),  
beside=TRUE, legend=TRUE)
```

解答類型分類率の算出



解答類型分類率の算出

- 設問1bの各条件，設問5の残りの条件（B条件，C条件）についても同様の手順で棒グラフの作成

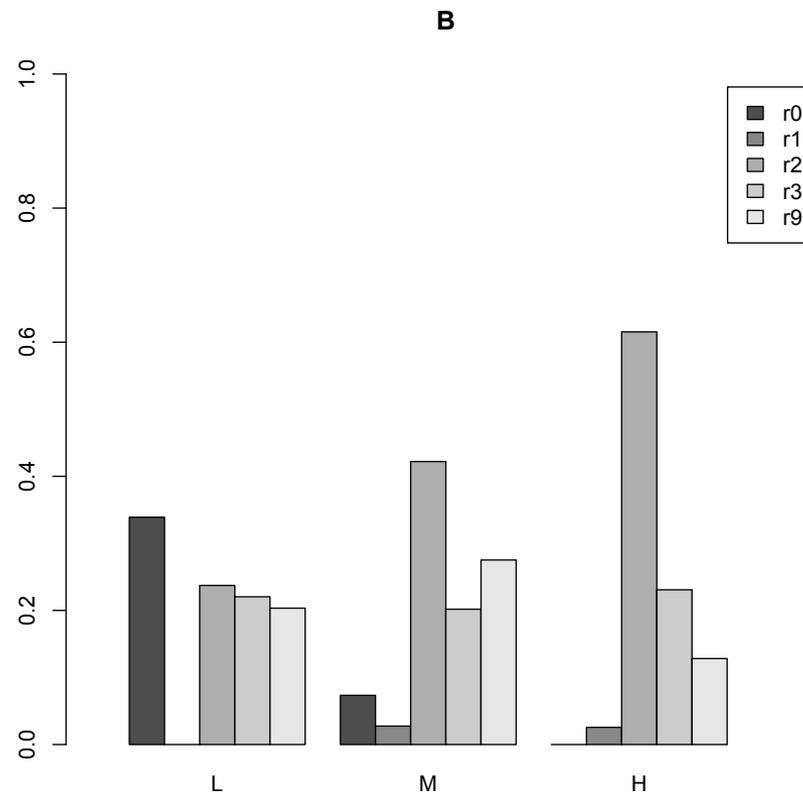
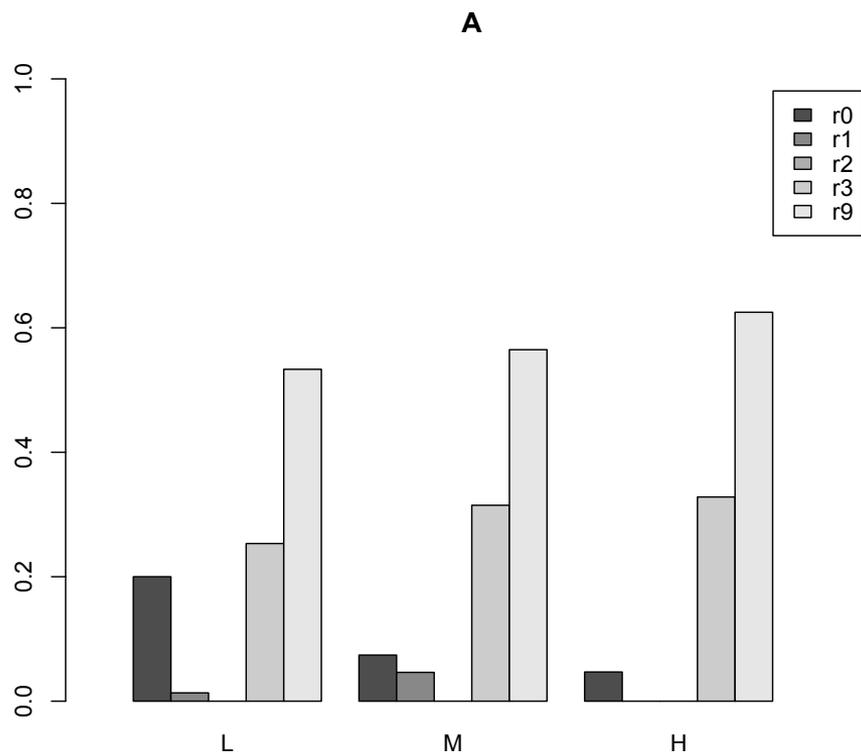
```
tcA1 <- table(A1$d1b, A1$gunwake) #設問1bのA条件の群と類型のクロス集計表
tcB1 <- table(B1$d1b, B1$gunwake) #設問1bのB条件の群と類型のクロス集計表
tcB5 <- table(B5$d5, B5$gunwake) #設問5のB条件の群と類型のクロス集計表
tcC5 <- table(C5$d5, C5$gunwake) #設問5のC条件の群と類型のクロス集計表

p.tcA1 <- prop.table(tcA1, 2) #設問1bのA条件における割合のクロス集計表を格納
p.tcB1 <- prop.table(tcB1, 2) #設問1bのB条件における割合のクロス集計表を格納
p.tcB5 <- prop.table(tcB5, 2) #設問5のB条件における割合のクロス集計表を格納
p.tcC5 <- prop.table(tcC5, 2) #設問5のC条件における割合のクロス集計表を格納

barplot(p.tcA1, main="A", xlim=c(0,20), ylim=c(0,1), beside=TRUE, legend=TRUE) #設問1bのA条件の棒グラフ
barplot(p.tcB1, main="B", xlim=c(0,20), ylim=c(0,1), beside=TRUE, legend=TRUE) #設問1bのB条件の棒グラフ
barplot(p.tcB5, main="B", xlim=c(0,20), ylim=c(0,1), beside=TRUE, legend=TRUE) #設問5のB条件の棒グラフ
barplot(p.tcC5, main="C", xlim=c(0,20), ylim=c(0,1), beside=TRUE, legend=TRUE) #設問5のC条件の棒グラフ
```

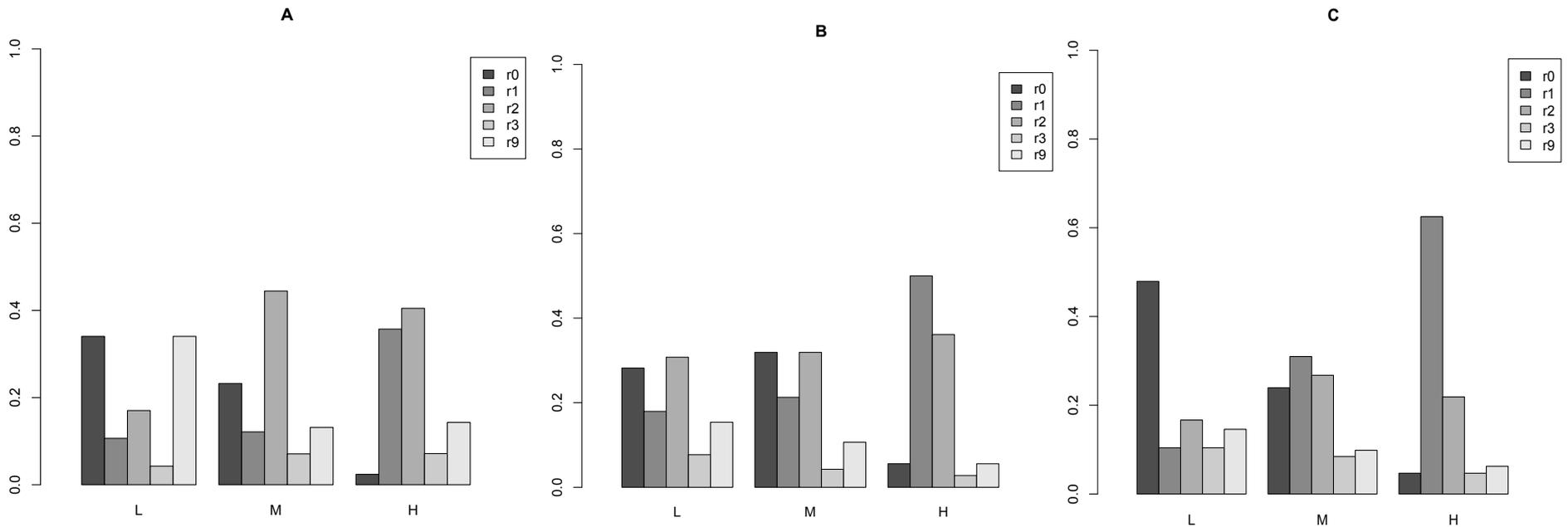
* `par(mfrow=c())`を用いて複数の図をまとめて表示することもできる

グラフ(設問1b)



設問1bにおける群ごとの解答分類率

グラフ(設問5)



設問5における群ごとの解答分類率

Overview

- 古典的テスト理論の解説 [宮坂]
- 安永ら(2012)の解説 [柘田]
- Rを用いた分析の実習
 - 合計得点の算出 [柘田]
 - 合計得点に基づいた群分け [柘田]
 - 古典的テスト理論における項目分析を用いた値の算出
 - 項目難易度 (項目得点率) の算出 [柘田]
 - 解答類型分類率の算出 [柘田]
 - 項目識別力(I-T相関)の算出 [宮坂]**
 - 得点率の差およびI-T相関の差に関する統計的推定 [宮坂]
 - 補足 [宮坂]
 - まとめ [宮坂]

項目識別力

- **項目識別力 (=項目弁別力, item discrimination)**
 - その項目がある特性（例：国語の能力）の高い人と低い人を区別できるかどうかの指標

項目識別力

- **項目弁別力指数 (item discrimination power index: DISC)**

- 解答が 0, 1 の場合, $-1 \sim +1$ の値をとる

- +1 に近い : 合計得点が高くなるにつれてその項目の得点率が高くなる。統計得点が低くなるにつれてその項目の得点率が低くなる。
 - 0 に近い : 合計得点の高低とその項目の得点率は関連しない。

- 目安

- 0.40以上 : とてもよい項目

- 0.30~0.39 : よい項目だが改良が必要かもしれない

- 0.20~0.29 : 改良が必要な項目

- 0.19以下 : 削除するか作り直す必要がある項目

項目識別力の種類

- **上位下位項目弁別指数 (upper-lower item discrimination index: U-L指数)**
 - その項目で上位群と下位群にどれほどの差がでるかを調べる
 - $(\text{合計得点上位27\%のうちその項目に正答した人数} - \text{合計得点下位27\%のうちその項目に正答した人数}) \div 27\% \text{の人数}$ (Johnson, 1951)
- **点双列相関係数 (point-biserial correlation coefficient)**

Item score (項目得点)と Total score (合計得点)の相関

 - I-T相関 (Item-Total correlation)
設問の得点と合計得点の相関
 - I-R相関 (Item-Remainder correlation)
設問の得点と残りの問題の合計得点の相関

項目識別力の種類

- 安永ら(2012)ではI-T相関を使用。
- 以下4つの設問を操作している。各設問の解答形式が参加者間で異なるため、これらの項目を除いた11項目を合計得点として使用。

問1b 「読解プロセス」

問5 「回答欄の字数制限」

問7a 「空所の表記法」

問8 「一文抜き出し問題」

識別力 (I-T相関) の算出手順

①点双列相関係数 (The point biserial coefficient of correlation) の算出 (Lew, 1949)

x は名義尺度・順序尺度, y は順序尺度・比率尺度

x : 1 か 0, y_i : $i = 1, \dots, n$

$x = 1$ の時 : $y = y_{1i}$ ($i = 1, \dots, n_1$ をとる), $M = \overline{y_1}$

$x = 0$ の時 : $y = y_{0i}$ ($i = 1, \dots, n_0$ をとる), $M = \overline{y_0}$

n : $n_1 + n_0$, $M = \overline{y}$, SD : y_i の SD , r : x_i と y_i の相関

$$r = \frac{\sqrt{\frac{n_1 n_0}{n}} (\overline{y_1} - \overline{y_0})}{\left[\sum_{i=0}^1 \sum_{j=1}^{n_i} (y_{ij} - \overline{y})^2 \right]^{1/2}}$$

=ピアソンの積率相関係数と同じ値になる

識別力 (I-T相関) の算出手順

- ピアソンの積率相関係数の算出
2変量(x, y)の共分散÷それぞれの標準偏差

$$r = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}}}$$
$$= \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})}{\sqrt{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2} \sqrt{(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \dots + (y_n - \bar{y})^2}}$$

識別力 (I-T相関) の算出手順

②母相関 (ρ) の検定

- 標本相関を t 値に変換する
- 帰無仮説 (H_0) : 母相関 (ρ) = 0

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

③ t 値から p 値を算出する

Rによる識別力(I-T相関)の算出

- **cor.test()**

- ピアソンの積率相関係数を算出する
- デフォルトで算出される値

t 値, 自由度, p 値, 95%信頼区間, 相関係数(r)

※相関係数の t 分布が $df=n-2$ の t 分布に従うことを利用して,
「2変量は無相関である」という帰無仮説を検討する

- 記述の仕方

cor.test(データセット 1, データセット 2)

- データセット 1: 各問い
- データセット 2: 合計得点(1b, 5, 7a, 8を除外したもの)

Rによる識別力(I-T相関)の算出

I-T相関の算出結果を別の変数に代入する

“A1”というデータセット(問1bがA条件であった参加者のデータのまとまり)の中の問題1bの正否(“s1b”)と合計得点(“goukei”)から相関係数を算出する

```
IT_A1b <- cor.test(A1$s1b,A1$goukei)
```

以下、同様に問1bのB条件, 問5のA条件, B条件, C条件

```
IT_B1b <- cor.test(B1$s1b,B1$goukei) #問1bのB条件
```

```
IT_A5 <- cor.test(A5$s5,A5$goukei) #問5のA条件
```

```
IT_B5 <- cor.test(B5$s5,B5$goukei) #問5のB条件
```

```
IT_C5 <- cor.test(C5$s5,C5$goukei) #問5のC条件
```

Rによる識別力(I-T相関)の算出

99%信頼区間を出したい場合は引数に**conf.level=0.99**を併記する。

cor.test(データセット1, データセット2, conf.level=0.99)

各結果の表示

IT_A1b #問1bのA条件の結果

IT_B1b #問1bのB条件の結果

IT_A5 #問5のA条件の結果

IT_B5 #問5のB条件の結果

IT_C5 #問5のC条件の結果

Rによる識別力(I-T相関)の算出

- ピアソンの積率相関係数
 - Rの出力結果 (例: 問1bのA条件)

```
> IT_A1b #問1bのA条件の結果
```

```
Pearson's product-moment correlation
```

```
data: A1$s1b and A1$goukei
```

```
t = 1.3903, df = 245, p-value = 0.1657
```

*t*値, 自由度, *p*値

```
alternative hypothesis: true correlation is not equal to 0
```

```
95 percent confidence interval:
```

95%信頼区間

```
-0.03675113 0.21096396
```

```
sample estimates:
```

```
cor
```

```
0.08847421
```

相関係数(I-T相関)

Rによる識別力 (I-T相関) の算出

- 問1bの結果一覧

	A条件	B条件
n	247	246
t 値	1.390	6.551
df	245	244
p 値	0.166	< .001
95%信頼区間	[-.037, .211]	[.275, .488]
99%信頼区間	[-.076, .248]	[.238, .518]
相関係数 r	.088	.387

Rによる識別力 (I-T相関) の算出

- 問5の結果一覧

	A条件	B条件	C条件
n	188	122	183
t 値	6.270	3.645	7.726
df	186	120	181
p 値	< .001	< .001	< .001
95%信頼区間	[.292, .529]	[.146, .477]	[.380, .600]
99%信頼区間	[.250, .561]	[.091, .510]	[.340, .628]
相関係数 r	.418	.316	.498

Rによる識別力 (I-T相関) の算出

特定の指標だけを表示させることもできる

IT_A1b\$estimate # 問1bのA条件の相関係数の推定値

IT_A1b\$conf.int # 問1bのA条件の信頼区間

IT_A1b\$p.value # 問1bのA条件の p 値

IT_B1b\$estimate # 問1bのB条件の相関係数の推定値

IT_B1b\$conf.int # 問1bのB条件の信頼区間

IT_B1b\$p.value # 問1bのB条件の p 値

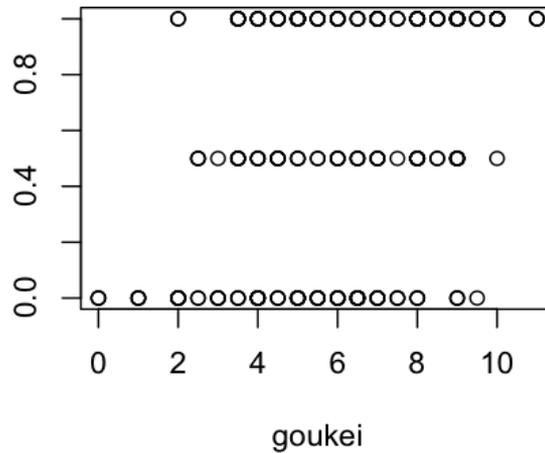
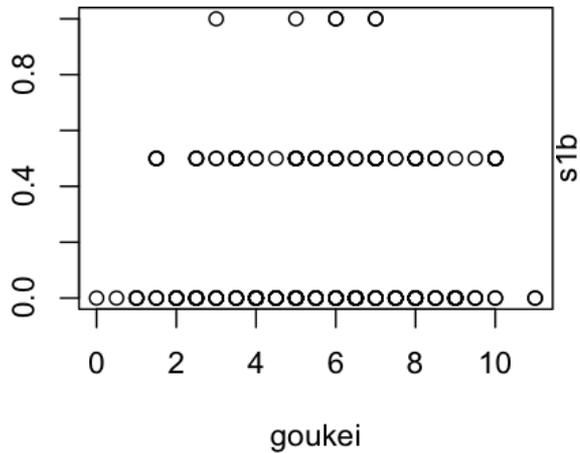
Rによる識別力(I-T相関)の算出

IT_A5\$estimate	# 問5のA条件の相関係数の推定値
IT_A5\$conf.int	# 問5のA条件の信頼区間
IT_A5\$p.value	# 問5のA条件の <i>p</i> 値
IT_B5\$estimate	# 問5のB条件の相関係数の推定値
IT_B5\$conf.int	# 問5のB条件の信頼区間
IT_B5\$p.value	# 問5のB条件の <i>p</i> 値
IT_C5\$estimate	# 問5のC条件の相関係数の推定値
IT_C5\$conf.int	# 問5のC条件の信頼区間
IT_C5\$p.value	# 問5のC条件の <i>p</i> 値

得点散布図を出してみた

1b-A : 具体例なし

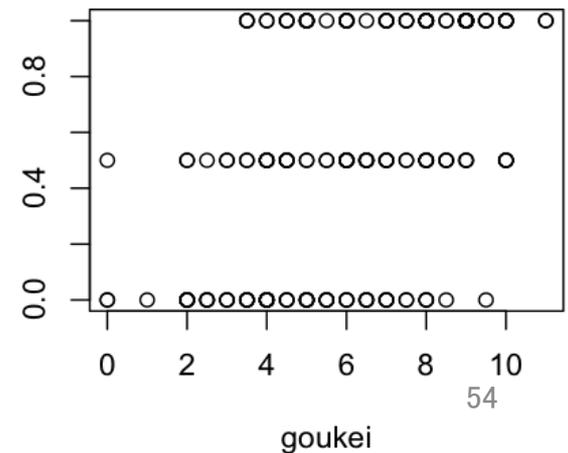
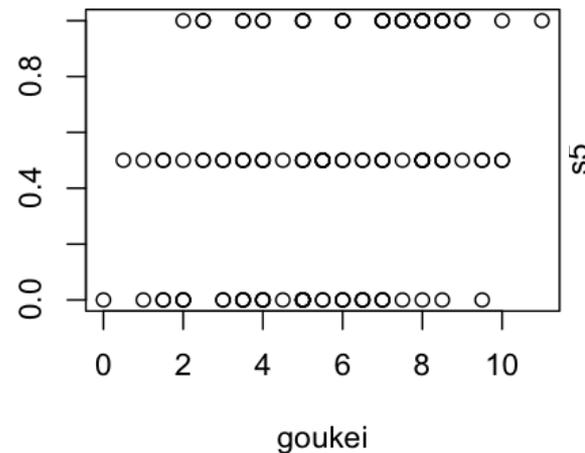
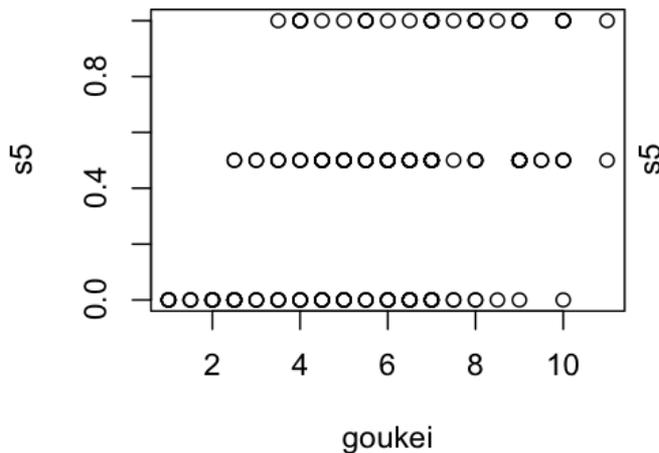
1b-B : 具体例あり



5-A : 45~55字

5-B : ~55字

5-C : 制限なし



Overview

- 古典的テスト理論の解説 [宮坂]
- 安永ら(2012)の解説 [柘田]
- Rを用いた分析の実習
 - 合計得点の算出 [柘田]
 - 合計得点に基づいた群分け [柘田]
 - 古典的テスト理論における項目分析を用いた値の算出
 - 項目難易度 (項目得点率) の算出 [柘田]
 - 回答累計分類率の算出 [柘田]
 - 項目識別力(I-T相関)の算出 [宮坂]
 - **得点率の差およびI-T相関の差に関する統計的推定** [宮坂]
 - 補足 [宮坂]
 - まとめ [宮坂]

得点率(比率)の差に関する推定の概要

1. 2標本の比率の差を出す
2. 信頼区間を出す
 - 臨界値(z値)を求める
 - 比率の差の標準誤差 (SE) を出す

$$SE = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

- z値と SE から 信頼区間を出す

$$CI = \text{標本平均の差} \pm \text{臨界値}(z\text{値}) \times \text{差の} SE$$

Rで得点率の差とその信頼区間を求める

```
p.dif <- function(pA,nA,pB,nB,qcrit=0.025){  
  p.difference <- pB-pA  
  seAB <- sqrt(((pA*(1-pA))/nA)+((pB*(1-pB))/nB))  
  q <- qnorm(qcrit,lower.tail=FALSE)  
  CIAB.L <- p.difference-q*seAB  
  CIAB.U <- p.difference+q*seAB  
  out <- cbind(p.difference, CIAB.L, CIAB.U)  
  return(out)  
}
```

入力しても変化なし・・・このスクリプトの意味は？

→次スライド

Rで得点率の差とその信頼区間を求める

- **function()**
 - ()内に引数, {}内に処理方法を記述し, 新たに関数を作成する
- 作りたいもの
 - (pA, nA, pB, nB)を引数とする“p.dif()”
 - A条件の得点率(pA), A条件の人数(nA), B条件の得点率(pB), B条件の人数(nB)を入力すると,
正答率の差, 95%信頼区間の下限と上限
を返す関数

Rで得点率の差とその信頼区間を求める

- 作る関数 `p.dif()` と引数の指定

- ()内に引数と“`qcrit=0.025`”を入力

- `qcrit`でスチューデント化された範囲分布(q)の臨界値(`qcritical`値)のデフォルトを指定する
- 95%信頼区間を出すため、今回は`qcrit=0.025` (両側検定のため $0.05/2$)とする

```
p.dif <- function(pA,nA,pB,nB,qcrit=0.025){
```

以下、引数を使って行う操作を指定

Rで得点率の差とその信頼区間を求める

- 正答率の差の算出
 - 2変量の正答率の差を“p.difference”に代入

```
p.difference <- pB-pA
```

Rで得点率の差とその信頼区間を求める

- 信頼区間(Confidence Interval: CI)の算出

- 標準誤差(SE)を“seAB”に代入

```
seAB <- sqrt(((pA*(1-pA))/nA)+((pB*(1-pB))/nB))
```

- 臨界値(z 値)を“q”に代入

- `qnorm(確率, lower.tail=FALSE)` : 標準正規分布上で指定した上側確率 p に対応する z 値を出す

- 確率 : 今回は臨界値(デフォルト `qcrit=0.025`)が自動的に代入されるように指定する

- `lower.tail` : `FALSE`は上側確率, `TRUE`は下側確率

```
q <- qnorm(qcrit,lower.tail=FALSE)
```

Rで得点率の差とその信頼区間を求める

- 信頼区間の算出（続き）

- 平均差の95%信頼区間

- 信頼区間= 標本平均の差 \pm 臨界値(z値) \times 差の標本標準誤差
- 下限を“CIAB.L”に，上限を“CIAB.U”に代入する

```
CIAB.L <- p.difference-q*seAB
```

```
CIAB.U <- p.difference+q*seAB
```

Rで得点率の差とその信頼区間を求める

- 結果の表示

- 行列を結合する

- `cbind()`

- 正答率の差(`p.difference`), 95%信頼区間の下限, 上限を横並びに“`out`”に代入する

```
out <- cbind(p.difference, CIAB.L, CIAB.U)
```

- “`out`”を表示する

```
return(out)
```

- 指定終わり

```
}
```

Rで得点率の差とその信頼区間を求める

`p.dif(pA,nA,pB,nB)`を使って、得点率と人数を入力

問1b：臨界値はデフォルト

`p.dif_1bAB<-p.dif(0.174, 247, 0.567, 246)` #A条件とB条件

問5：検定を繰り返すため、type-I errorを全体で $\alpha = 0.05$ に抑えるBonferroniの方法を使用。 $0.05 * 1/3(=組合せ) * 1/2(=上側確率を算出) = 0.008$

`p.dif_5AB<-p.dif(0.386, 188, 0.475, 122, 0.008)` #A条件とB条件

`p.dif_5BC<-p.dif(0.475, 122, 0.516, 183, 0.008)` #B条件とC条件

`p.dif_5CA<-p.dif(0.516, 183, 0.386, 188, 0.008)` #C条件とA条件

Rで得点率の差とその信頼区間を求める

問1bの結果

p.dif_1bAB

問5の結果

p.dif_5AB

p.dif_5BC

p.dif_5CA

```
> # 問1bの結果
> p.dif_1bAB
      p.difference      CIAB.L      CIAB.U
[1,]          0.393 0.3150957 0.4709043
>
> # 問5の結果
> p.dif_5AB
      p.difference      CIAB.L      CIAB.U
[1,]          0.089 -0.04948053 0.2274805
> p.dif_5BC
      p.difference      CIAB.L      CIAB.U
[1,]          0.041 -0.09964392 0.1816439
> p.dif_5CA
      p.difference      CIAB.L      CIAB.U
[1,]         -0.13 -0.2534291 -0.006570851
```

↑得点率の差

↑信頼区間

I-T相関(相関係数)の信頼区間

1. 相関係数 r をFisherの z 変換 (逆双曲線正接関数)

$$z_1 = \frac{1}{2} \log \frac{1+r_1}{1-r_1}, \quad z_2 = \frac{1}{2} \log \frac{1+r_2}{1-r_2}$$

2. I-T相関の差の標準誤差 (SE) を出す ($z_A - z_B$ の SE)

$$SE = \sqrt{SE_A^2 + SE_B^2} = \sqrt{\frac{1}{n_A - 3} + \frac{1}{n_B - 3}}$$

3. z 値と SE から信頼区間を出す ($CI = z_1 - z_2 \pm z_0 \cdot SE$)
4. z 値を逆変換して r に戻す(下記を下限と上限それぞれに行う)

自然対数の底の(2×下限or上限)乗-1

÷自然対数の底の(2×下限or上限)乗+1

I-T相関の差に関する統計的推定

```
r.dif <- function(rA,nA,rB,nB,qcrit=0.025){  
r.difference <- rB-rA  
zA=1/2*log((1+rA)/(1-rA))  
zB=1/2*log((1+rB)/(1-rB))  
seAB <- sqrt((1/(nA-3))+(1/(nB-3)))  
q <- qnorm(qcrit,lower.tail=FALSE)  
zAB.L <- zB-zA-q*seAB  
zAB.U <- zB-zA+q*seAB  
CIAB.L <- ((exp(2*zAB.L))-1)/((exp(2*zAB.L))+1)  
CIAB.U <- ((exp(2*zAB.U))-1)/((exp(2*zAB.U))+1)  
out <- cbind(r.difference, CIAB.L, CIAB.U)  
return(out)  
}
```

RでI-T相関の差とその信頼区間を求める

- **function()**

- ()内に引数, {}内に処理方法を記述し, 新たに関数を作成する

- 作りたいもの

- (rA, nA, rB, nB)を引数とする“r.dif()”

- A条件のI-T相関(rA), A条件の人数(nA), B条件のI-T相関(rB), B条件の人数(nB)を入力すると,

- I-T相関の差, 95%信頼区間の下限と上限**
を返す関数

RでI-T相関の差とその信頼区間を求める

- 作る関数 `r.dif()` と引数の指定
 - ()内に引数と“`qcrit=0.025`”を入力
 - `qcrit`でスチューデント化された範囲分布(q)の臨界値(`qcritical`値)のデフォルトを指定する
 - 95%信頼区間を出すため、今回は`qcrit=0.025` (両側検定のため $0.05/2$)とする

```
r.dif <- function(rA,nA,rB,nB,qcrit=0.025){
```

以下、引数を使って行う操作を指定

RでI-T相関の差とその信頼区間を求める

- I-T相関の差の算出
 - 2変量のI-T相関の差を“r.difference”に代入

```
r.difference <- rB-rA
```

RでI-T相関の差とその信頼区間を求める

- 信頼区間(Confidence Interval: CI)の算出

- 相関係数 r を z 変換

$$z_A = 1/2 * \log((1+r_A)/(1-r_A))$$

$$z_B = 1/2 * \log((1+r_B)/(1-r_B))$$

- 標準誤差(SE)を“ se_{AB} ”に代入

$$se_{AB} <- \sqrt{(1/(n_A-3)) + (1/(n_B-3))}$$

RでI-T相関の差とその信頼区間を求める

- 信頼区間の算出（続き）

- 臨界値(z 値)を“ q ”に代入

- `qnorm(確率, lower.tail=FALSE)`：標準正規分布上で指定した上側確率 p に対応する z 値を出す

- 確率：今回は臨界値(デフォルト `qcrit=0.025`)が自動的に代入されるように指定する

- `lower.tail`：FALSEは上側確率，TRUEは下側確率

```
q <- qnorm(qcrit,lower.tail=FALSE)
```

RでI-T相関の差とその信頼区間を求める

- 信頼区間の算出 (続き)

- 平均差の95%信頼区間

- 信頼区間= 標本平均の差 \pm 臨界値(z 値) \times 差の標本標準誤差
- 下限を“ $z_{AB.L}$ ”に, 上限を“ $z_{AB.U}$ ”に代入する

$$z_{AB.L} <- z_B - z_A - q * se_{AB}$$

$$z_{AB.U} <- z_B - z_A + q * se_{AB}$$

- z を r に戻す

$$CI_{AB.L} <- ((\exp(2 * z_{AB.L})) - 1) / ((\exp(2 * z_{AB.L})) + 1)$$

$$CI_{AB.U} <- ((\exp(2 * z_{AB.U})) - 1) / ((\exp(2 * z_{AB.U})) + 1)$$

RでI-T相関の差とその信頼区間を求める

- 結果の表示

- 行列を結合する

- `cbind()`

- 正答率の差(`p.difference`), 95%信頼区間の下限, 上限を横並びに“`out`”に代入する

```
out <- cbind(r.difference, CIAB.L, CIAB.U)
```

- “`out`”を表示する

```
return(out)
```

- 指定終わり

```
}
```

RでI-T相関の差とその信頼区間を求める

`r.dif(pA,nA,pB,nB)`を使って，得点率と人数を入力

問1b：臨界値はデフォルト

`r.dif_1bAB<-r.dif(0.088, 247, 0.387, 246)` #A条件とB条件

問5：検定を繰り返すため，type-I errorを全体で $\alpha = 0.05$ に抑えるBonferroniの方法を使用。 $0.05 * 1/3(=組合せ) * 1/2(=上側確率を算出) = 0.008$

`r.dif_5AB<-r.dif(0.418, 188, 0.316, 122, 0.008)` #A条件とB条件

`r.dif_5BC<-r.dif(0.316, 122, 0.498, 183, 0.008)` #B条件とC条件

`r.dif_5CA<-r.dif(0.498, 183, 0.418, 188, 0.008)` #C条件とA条件

Rで得点率の差とその信頼区間を求める

問1bの結果

r.dif_1bAB

問5の結果

r.dif_5AB

r.dif_5BC

r.dif_5CA

```
> # 問1bの結果
> r.dif_1bAB
      r.difference      CIAB.L      CIAB.U
[1,]          0.299  0.1414543  0.460281
>
> # 問5の結果
> r.dif_5AB
      r.difference      CIAB.L      CIAB.U
[1,]         -0.102 -0.380926  0.163523
> r.dif_5BC
      r.difference      CIAB.L      CIAB.U
[1,]          0.182 -0.06507022  0.4652995
> r.dif_5CA
      r.difference      CIAB.L      CIAB.U
[1,]         -0.08 -0.3395443  0.1496903
```

↑I-T相関の差

↑信頼区間

結果の理解

差の計算式 = (後に入力した群 - 前に入力した群)
と作った

• 問1b : B-Aの結果

A条件の得点率 : 0.174, B条件の得点率 : 0.567

得点率の差 : .393 [.315, .417]

→ B条件 > A条件

A条件のI-T相関 : 0.088, B条件のI-T相関 : 0.387

I-T相関の差 : .299 [.141, .460]

→ B条件 > A条件

結果の理解

- 問5：B-Aの結果 信頼区間が0をまたいでいる...

A条件の得点率：0.386, B条件の得点率0.475

得点率の差 .089 [-.049, .227]

→B-Aが-5%になる(A条件が5%高い)ことも

23%になる(B条件が23%高い)こともあり得る

A条件のI-T相関：0.418, B条件のI-T相関：0.316

I-T相関の差 -.102 [-.381, .164]

→B-Aが-38%になる(A条件が38%高い)ことも

16%になる(B条件が16%高い)こともあり得る

結果の理解

- 問5：C-Bの結果 信頼区間が0をまたいでいる...

B条件の得点率：0.475, C条件の得点率：0.516

得点率の差：.041 [-.100, .182]

→C-Bが-10%になる(B条件が10%高い)ことも
18%になる(C条件が18%高い)こともあり得る

B条件のI-T相関：0.316, C条件のI-T相関：0.498

I-T相関の差：.182 [-.065, .465]

→C-Bが-6%になる(B条件が6%高い)ことも
47%になる(C条件が47%高い)こともあり得る

結果の理解

- 問5：A-Cの結果 I-T相関は0をまたいでいる

C条件の得点率：0.516, A条件の得点率：0.386

得点率の差 $-.130$ [**-.253**, **-.007**]

→A-Cが**-25%**になる(C条件が25%高い)ことも

-0.7%になる(C条件が0.7%高い)こともあり得る

C条件I-T相関：0.498, A条件I-T相関：0.418

I-T相関の差 $-.080$ [**-.340**, **.150**]

→A-Cが**-34%**になる(C条件が34%高い)ことも

15%になる(A条件が15%高い)こともあり得る

自作スクリプトの保存と使用

自作のスクリプトはメモ帳などのテキストファイルに保存して読み込み，使用することができる。

1. メモ帳などのテキストファイルに記述
2. 拡張子を“.R”として(例えば“rdif.R”)作業ディレクトリ内に保存
3. `source(“rdif.R”)`を用いて読み込む

Overview

- 古典的テスト理論の解説 [宮坂]
- 安永ら(2012)の解説 [柘田]
- Rを用いた分析の実習
 - 合計得点の算出 [柘田]
 - 合計得点に基づいた群分け [柘田]
 - 古典的テスト理論における項目分析を用いた値の算出
 - 項目難易度（項目得点率）の算出 [柘田]
 - 回答累計分類率の算出 [柘田]
 - 項目識別力(I-T相関)の算出 [宮坂]
 - 得点率の差およびI-T相関の差に関する統計的推定 [宮坂]
 - **補足** [宮坂]
 - まとめ [宮坂]

U-L指数の算出

U-L指数

Brennan (1972)による改訂版(The Discrimination Index B)

= 上位の正答者数÷上位の人数

- 下位の正答者数÷下位の人数

= 上位27%の正答率-下位27%の正答率

今回は条件差までは比較しないので，合計点として全13項目を使いたいと思います

そこで，全13項目の合計点を“goukei2”として追加

```
saiten$goukei2 <- with(saiten, s1a+s1b+s2a+s2b+s3+s4+s5+s6a+s6b+s6c+s7a+s7b+s7c+s8+s9)
```

U-L指数の算出

各設問の各条件について27%の人数を算出する

$$247 * 0.27 \quad \# \text{ 問1bのA条件} = 66.69$$

$$246 * 0.27 \quad \# \text{ 問1bのB条件} = 66.42$$

$$188 * 0.27 \quad \# \text{ 問5のA条件} = 50.76$$

$$122 * 0.27 \quad \# \text{ 問5のB条件} = 32.94$$

$$183 * 0.27 \quad \# \text{ 問5のC条件} = 49.41$$

U-L指数の算出

問1bのA条件

```
A1_2 <- subset(saiten, e1=="A", c(id, e1, d1b, s1b,  
goukei2))
```

```
A1_order <- order(A1_2$goukei2, A1_2$id,  
decreasing=TRUE)
```

```
df_A1_order<-A1_2[A1_order,]
```

```
A1_H<-head(df_A1_order,n=66.69)
```

```
summary(A1_H)
```

```
A1_L<-tail(df_A1_order,n=66.69)
```

```
summary(A1_L)
```

```
UL_A1<-mean(A1_H$s1b)-mean(A1_L$s1b)
```

```
# →スクリプトの意味は次スライドより
```

U-L指数の算出

- 下準備

- 全13項目の合計点を“goukei2”として“saiten”に追加

```
saiten$goukei2 <- with(saiten, s1a+s1b+s2a+s2b+s3+s4+s5+s6a+s6b  
+s6c+s7a+s7b+s7c+s8+s9)
```

- “saiten”から問1aのA条件（うち, id, e1, d1b, s1b, goukei2）を抽出し, “A1_2”というデータセットを作る

```
A1_2 <- subset(saiten, e1=="A", c(id, e1, d1b, s1b, goukei2))
```

U-L指数の算出

- 元の“saiten”

```
> head(saiten)
  id e1 e5 d1a d1b d2a d2b d3 d4 d5 d6a d6b d6c d7a d7b d7c d8 d9 s1a s1b s2a s2b s3 s4 s5 s6a s6b s6c s7a s7b s7c s8 s9
1  1  B  C  r2  r9  r1  r5  r2  r1  r2  r3  r1  r4  r9  r2  r2  r1  r2  1.0  0.0  0  0  1  1  0.5  0  1  1  0  0  1  1  0
2  2  B  C  r2  r2  r2  r5  r2  r1  r1  r7  r1  r4  r1  r9  r1  r3  r1  1.0  1.0  1  0  1  1  1.0  0  1  1  1  0  1  1  1
3  3  B  C  r3  r9  r4  r3  r2  r9  r0  r6  r4  r3  r9  r1  r1  r9  r4  0.5  0.0  0  0  1  0  0.0  1  0  0  0  1  1  0  0
4  4  B  C  r2  r3  r2  r5  r2  r1  r1  r6  r7  r4  r0  r2  r0  r0  r3  1.0  0.5  1  0  1  1  1.0  1  0  1  0  0  0  0  0  0
5  5  B  C  r3  r2  r2  r5  r2  r9  r2  r6  r1  r4  r9  r9  r9  r9  r1  0.5  1.0  1  0  1  0  0.5  1  1  1  0  0  0  0  1
6  6  B  C  r2  r3  r2  r5  r2  r1  r1  r6  r1  r7  r1  r1  r2  r1  r4  1.0  0.5  1  0  1  1  1.0  1  1  0  1  1  1  1  0
```

- “goukei”を追加

```
> head(saiten)
  id e1 e5 d1a d1b d2a d2b d3 d4 d5 d6a d6b d6c d7a d7b d7c d8 d9 s1a s1b s2a s2b s3 s4 s5 s6a s6b s6c s7a s7b s7c s8 s9 goukei gunwake
1  1  B  C  r2  r9  r1  r5  r2  r1  r2  r3  r1  r4  r9  r2  r2  r1  r2  1.0  0.0  0  0  1  1  0.5  0  1  1  0  0  1  1  0  6.0  M
2  2  B  C  r2  r2  r2  r5  r2  r1  r1  r7  r1  r4  r1  r9  r1  r3  r1  1.0  1.0  1  0  1  1  1.0  0  1  1  1  0  1  1  1  8.0  H
3  3  B  C  r3  r9  r4  r3  r2  r9  r0  r6  r4  r3  r9  r1  r1  r9  r4  0.5  0.0  0  0  1  0  0.0  1  0  0  0  1  1  0  0  4.5  M
4  4  B  C  r2  r3  r2  r5  r2  r1  r1  r6  r7  r4  r0  r2  r0  r0  r3  1.0  0.5  1  0  1  1  1.0  1  0  1  0  0  0  0  0  6.0  M
5  5  B  C  r3  r2  r2  r5  r2  r9  r2  r6  r1  r4  r9  r9  r9  r9  r1  0.5  1.0  1  0  1  0  0.5  1  1  1  0  0  0  0  1  6.5  M
6  6  B  C  r2  r3  r2  r5  r2  r1  r1  r6  r1  r7  r1  r1  r2  r1  r4  1.0  0.5  1  0  1  1  1.0  1  1  0  1  1  1  1  0  8.0  H
```

- “goukei2”を追加

```
> head(saiten)
  id e1 e5 d1a d1b d2a d2b d3 d4 d5 d6a d6b d6c d7a d7b d7c d8 d9 s1a s1b s2a s2b s3 s4 s5 s6a s6b s6c s7a s7b s7c s8 s9 goukei gunwake goukei2
1  1  B  C  r2  r9  r1  r5  r2  r1  r2  r3  r1  r4  r9  r2  r2  r1  r2  1.0  0.0  0  0  1  1  0.5  0  1  1  0  0  1  1  0  6.0  M  7.5
2  2  B  C  r2  r2  r2  r5  r2  r1  r1  r7  r1  r4  r1  r9  r1  r3  r1  1.0  1.0  1  0  1  1  1.0  0  1  1  1  0  1  1  1  8.0  H  12.0
3  3  B  C  r3  r9  r4  r3  r2  r9  r0  r6  r4  r3  r9  r1  r1  r9  r4  0.5  0.0  0  0  1  0  0.0  1  0  0  0  1  1  0  0  4.5  M  4.5
4  4  B  C  r2  r3  r2  r5  r2  r1  r1  r6  r7  r4  r0  r2  r0  r0  r3  1.0  0.5  1  0  1  1  1.0  1  0  1  0  0  0  0  0  6.0  M  7.5
5  5  B  C  r3  r2  r2  r5  r2  r9  r2  r6  r1  r4  r9  r9  r9  r9  r1  0.5  1.0  1  0  1  0  0.5  1  1  1  0  0  0  0  1  6.5  M  8.0
6  6  B  C  r2  r3  r2  r5  r2  r1  r1  r6  r1  r7  r1  r1  r2  r1  r4  1.0  0.5  1  0  1  1  1.0  1  1  0  1  1  1  1  0  8.0  H  11.5
```

U-L指数の算出

- 上位27%を抽出する
 - 「データセット“A1_2”(問1aのA条件を抽出したものを)を合計点の降順，同値の場合は“id”の降順に並べ替える」という順番の情報を作成し，“A1_order”に格納する

```
A1_order <- order(A1_2$goukei2, A1_2$id, decreasing=TRUE)
```

- 作った順番データ“A1_order”を使って“A1_2”を並べ替え，その結果を“df_A1_order”に格納する

```
df_A1_order <- A1_2[A1_order,]
```

U-L指数の算出

- “df_A1_order”から上位75%を抽出して“A1_H”に、
下位75%を抽出して“A1_L”に入れる

```
A1_H<-head(df_A1_order,n=66.69)
```

```
summary(A1_H) # 記述統計量の確認
```

```
A1_L<-tail(df_A1_order,n=66.69)
```

```
summary(A1_L) # 記述統計量の確認
```

- “A1_H”のs1bの平均値と“A1_L”のs1bの平均値の差
を“UL_A1”に代入する

```
UL_A1<-mean(A1_H$s1b)-mean(A1_L$s1b)
```

U-L指数の算出

問1bのB条件

```
B1_2 <- subset(saiten, e1=="B", c(id, e1, d1b, s1b, goukei2))
```

```
B1_order <- order(B1_2$goukei2, B1_2$id, decreasing=TRUE)
```

```
df_B1_order <- B1_2[B1_order,]
```

```
B1_H <- head(df_B1_order, n=66.42)
```

```
summary(B1_H)
```

```
B1_L <- tail(df_B1_order, n=66.42)
```

```
summary(B1_L)
```

```
UL_B1 <- mean(B1_H$s1b) - mean(B1_L$s1b)
```

U-L指数の算出

問5のA条件

```
A5_2 <- subset(saiten, e5=="A", c(id, e5, d5, s5, goukei2))
```

```
A5_order <- order(A5_2$goukei2, A5_2$id,  
decreasing=TRUE)
```

```
df_A5_order<-A5_2[A5_order,]
```

```
A5_H<-head(df_A5_order,n=50.76)
```

```
summary(A5_H)
```

```
A5_L<-tail(df_A5_order,n=50.76)
```

```
summary(A5_L)
```

```
UL_A5<-mean(A5_H$s5)-mean(A5_L$s5)
```

U-L指数の算出

問5のB条件

```
B5_2 <- subset(saiten, e5=="B", c(id, e5, d5, s5, goukei2))
```

```
B5_order <- order(B5_2$goukei2, B5_2$id,  
decreasing=TRUE)
```

```
df_B5_order<-B5_2[B5_order,]
```

```
B5_H<-head(df_B5_order,n=32.94)
```

```
summary(B5_H)
```

```
B5_L<-tail(df_B5_order,n=32.94)
```

```
summary(B5_L)
```

```
UL_B5<-mean(B5_H$s5)-mean(B5_L$s5)
```

U-L指数の算出

問5のC条件

```
C5_2 <- subset(saiten, e5=="C", c(id, e5, d5, s5, goukei2))
```

```
C5_order <- order(C5_2$goukei2, C5_2$id,  
decreasing=TRUE)
```

```
df_C5_order<-C5_2[C5_order,]
```

```
C5_H<-head(df_C5_order,n=49.41)
```

```
summary(C5_H)
```

```
C5_L<-tail(df_C5_order,n=49.41)
```

```
summary(C5_L)
```

```
UL_C5<-mean(C5_H$s5)-mean(C5_L$s5)
```

U-L指数の算出

結果の表示

UL_A1

UL_B1

UL_A5

UL_B5

UL_C5

合計点として
13項目全てを用いた場合

```
> UL_A1  
[1] 0.0928313  
> UL_B1  
[1] 0.5418363  
> UL_A5  
[1] 0.542549  
> UL_B5  
[1] 0.4469697  
> UL_C5  
[1] 0.6971429
```

操作した4項目を除外した
合計点を用いた場合

```
> UL_A1  
[1] 0.03245138  
> UL_B1  
[1] 0.3772049  
> UL_A5  
[1] 0.3245098  
> UL_B5  
[1] 0.3087121  
> UL_C5  
[1] 0.5559184
```

Overview

- 古典的テスト理論の解説 [宮坂]
- 安永ら(2012)の解説 [杵田]
- Rを用いた分析の実習
 - 合計得点の算出 [杵田]
 - 合計得点に基づいた群分け [杵田]
 - 古典的テスト理論における項目分析を用いた値の算出
 - 項目難易度 (項目得点率) の算出 [杵田]
 - 回答累計分類率の算出 [杵田]
 - 項目識別力(I-T相関)の算出 [宮坂]
 - 得点率の差およびI-T相関の差に関する統計的推定 [宮坂]
 - 補足 [宮坂]
 - **まとめ** [宮坂]

結果のまとめ

- 研究の目的（の一部）

「読解プロセス」と「回答欄の字数制限」が受検者の回答におよぼす影響の検討

結果のまとめ

- 「読解プロセス」
 - A条件：傍線部分の段落に具体例が載っていないパターン (統合・解釈)
 - B条件：具体例が載っているパターン (情報へのアクセス・取り出し)

結果のまとめ

問1bより

- 正答となる具体例がない条件[A]
 - 得点率が低い
 - 識別力も低い
 - 誤答の内容を記述する割合が多い
 - 自分で具体例を記述する割合は少ない
 - 「**統合・解釈**」が求められ、回答が困難になる
- 具体例が本文にある条件[B]
 - 得点率50%程度
 - 本文にある具体例を抜き出す割合が多い
 - 自分で具体例を記述する割合は少ない
 - 「**情報へのアクセス・取り出し**」はできる

結果のまとめ

テスト作成の平行項目（同類の項目）作成時には...

- 設問文やその構成などの表面的な側面を類似させるだけでなく，読解プロセスにまで踏み込んで同類の設問となるように作成することが求められる。

結果のまとめ

- 「回答欄の字数」
 - A条件：四十五字以上五十五字以内で書きなさい
 - B条件：五十五字以内で書きなさい
 - C条件：字数制限を行わない

結果のまとめ

問5より

- 得点率

字数制限なし条件[C]は...

– 55字以内条件[B]と同程度

– 45字以上55字以内条件[A]よりも高い

→難易度は「**字数制限なし**」より「**45字以上55字以内**」の方が高い

結果のまとめ

問5より

- 識別力

字数制限なし条件[C]は...

- 55字以内条件[B]よりも高い

- 45字以上55字以内条件[A]と同程度

→回答欄は「55字以内」よりも「字数制限なし」の方が合計得点の高い人と低い人をより良く区別する。

結果のまとめ

問5より

- 回答の内容

- 45字以上55字以内[A]

- 西洋の特徴と異なる内容のみ

- 55字以内で書きなさい[B]

- 西洋の特徴と異なる内容のみ

- 字数制限なし[C]

- 西洋の特徴と異なる内容と日本の特徴

→**字数制限がある場合，西洋の特徴と異なる内容を記述すると制限字数に達してしまい，日本の特徴に関する内容を記述できない**

結果のまとめ

本文や設問をどの程度理解しているかを知る目的で問題を設定している場合...

- 字数制限を設けないことが有効
- ただし合計得点の低い受検者の無効回答が多くなる

結果のまとめ

本研究の結果が示唆すること

- わずかな構造的性質の操作によって受検者の回答に変化が生じる
- 構造的性質について実証的に検討することの意義を示している

限界点

- 構造的性質に関して1つの題材（問題文）に基づいた結果である

今後の研究

- より多くの題材を用いた検証が必要（教育現場に還元できるものに焦点を当てる，原典の著作権に配慮する必要がある）

この研究についてひとこと

- 本研究著者の研究は「設問はこのように作られるべきである」ということを主張するもの？
 - No。
 - 設問形式において最も重要となるのは，作成者の測定意図。テスト作成者が測りたいものを測れるようになることをサポートするためのもの。
 - 将来的には，テスト作成者が測定意図と具体的な項目得点率や識別力の値とを照らし合わせながら，複数の選択肢から自らの望む設問設定を選択できる状況を提供したい。

References

[栞田]

- 服部環 (2011). 心理・教育のためのRによるデータ分析. 福村出版
- 加藤健太郎・山田剛史・川端一光 (2014). Rによる項目反応理論. オーム社
- 石原知英 (2014). 古典的テスト理論を用いた2012年度新入生英語プレイスメントテストの分析と改善への提言. 言語と文化：愛知大学語学教育研究室紀要, **57**, 1-10.
- 舟尾暢男(2009). The R tips –データ解析環境Rの基本技・グラフィック活用集 (PDF版) . オーム社

References

[宮坂]

- Brennan, L. Robert. (1972). A generalized upper-lower item discrimination index. *Educ. Psychol. Meas.*, 32, 289-303.
- 石井 秀宗 (2014). 人間科学のための統計分析 ころろに関心があるすべての人のために 医歯薬出版株式会社
- Johnson, A. Pemberton. (1951). Notes on a suggested index of item validity: The UL Index. *J. Educ. Psychol.*, 42 (8), 499-504.
- Lew Joseph. (1949). The point biserial coefficient of correlation. *Ann. Math. Stat*, 20, 125-126.
- 南風原 朝和 (2002). 心理統計学の基礎 総合的理解のために 有斐閣アルマ
- 内田 治・西澤 英子 (2012). Rによる統計的検定と推定 オーム社
- 渡部 倫子 (2012). 日本語プレースメントテストにおける文法テスト項目の改訂 広島大学大学院教育学研究科紀要, 239-244.