

(3)大学ごとの標準化

標準化の一つとしてz得点を求める。

```
> A_z <- (A-A_mean)/A_std
> A_z
[1] -0.6682061 -0.1542014 -0.7967072 -0.9252084
-0.7967072 1.5163138
[7] 0.1028009 1.6448149 1.1308103 -1.0537096

> B_z <- (B-B_mean)/B_std
> B_z
[1] -0.4731602 -0.1577201 -0.7886004 -0.4731602
1.1040405 0.4731602
[7] -1.1040405 -1.4194807 1.1040405 1.7349208
```

7

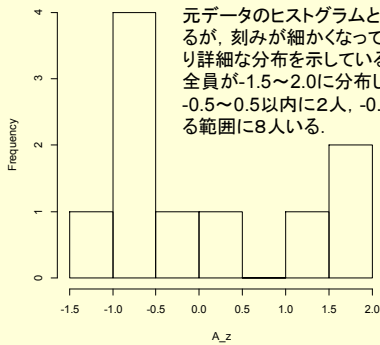
ヒストグラムの範囲と刻みの指定方法

```
hist(aaa, breaks=seq(10,100,5))
```

10~100の範囲で5刻みのヒストグラムを作図  
 10~100は全てのデータを含むように指定  
 (例) データ分布: 5~125  
 範囲指定: 5~125, 0~130など

10

Histogram of A\_z



元データのヒストグラムとほぼ同じであるが、刻みが細かくなっているため、より詳細な分布を示している。  
 全員が-1.5~2.0に分布している。  
 -0.5~0.5以内に2人、-0.5~0.5を超える範囲に8人いる。

8

第9章 データフレーム

Rで利用されるデータの保存形式  
 数値や文字など異なるタイプのデータを扱うことができる

◆データフレームの作成法  
 Excelで表を作成(数値, 文字混在)  
 csvファイルとして保存→aaa.csv  
 txtファイルとして保存→bbb.txt

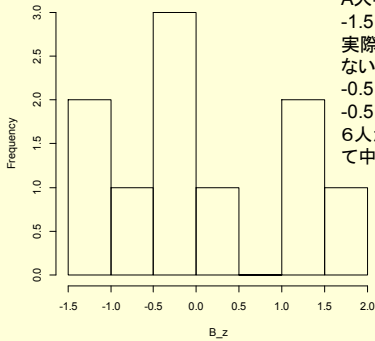
read.csv("aaa.csv") 表題(ヘッダー)があることが前提  
 read.csv("aaa.csv", header=FALSE) ヘッダーがない場合

read.table("aaa.csv") ヘッダーがないことが前提  
 read.table("aaa.csv", header=TRUE) ヘッダーがある場合

read.table(bbb.txt") txtファイルはread.table()を使用

14

Histogram of B\_z



A大学と同様に全員が-1.5~2.0に分布しており、実際の時間数に左右されない分布状況が分かる。  
 -0.5~0.5以内に4人、  
 -0.5~0.5を超える範囲に6人おり、A大学に比べて中央に分布している。

9

テキストファイル(table.txt)

No	名前	性別	数学	統計	心理テスト	統計テスト1	統計テスト2	指導法
1	大村	男	嫌い	好き	13	6	10	C
2	本多	男	嫌い	好き	14	10	13	B
3	杉内	女	好き	嫌い	6	6	14	A

Excelで編集して、txtファイルとして保存する→作業ディレクトリ(推奨する)

適当なエディタで編集して、txtファイルとして保存する→作業ディレクトリ  
 半角スペースで区切ること。全角スペースは文字と見なされる。  
 改行は認識される。

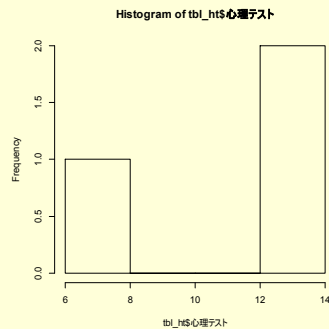
15

```
> tbl <- read.table("table.txt") ヘッダー無しが標準
> tbl
  V1 V2 V3 V4 V5 V6 V7 V8 V9
1 No. 名前 性別 数学 統計 心理テスト 統計テスト1 統計テスト2 指導法
2 1 大村 男 嫌い 好き 13 6 10 C
3 2 本多 男 嫌い 好き 14 10 13 B
4 3 杉内 女 好き 嫌い 6 6 14 A

> tbl_ht <- read.table("table.txt", header=TRUE) ヘッダーがある場合
> tbl_ht
No. 名前 性別 数学 統計 心理テスト 統計テスト1 統計テスト2 指導法
1 1 大村 男 嫌い 好き 13 6 10 C
2 2 本多 男 嫌い 好き 14 10 13 B
3 3 杉内 女 好き 嫌い 6 6 14 A
```

16

```
> hist(tbl_ht$心理テスト)
```



19

この頁以降の内容はcsvファイルを読み込んだ場合でも同じである。

```
> table(tbl_ht$指導法)
```

```
A B C
1 1 1
```

```
> table(tbl_ht[,9])
```

```
A B C
1 1 1
```

17

```
> mean(tbl_ht$統計テスト1)
[1] 7.333333
```

```
> var(tbl_ht$統計テスト2)
[1] 4.333333
```

```
> sd(tbl_ht$統計テスト1)
[1] 2.309401
```

```
> min(tbl_ht$統計テスト1)
[1] 6
```

```
> max(tbl_ht$統計テスト1)
[1] 10
```

20

```
> tbl_ht[2,]
No. 名前 性別 数学 統計 心理テスト 統計テスト1 統計テスト2 指導法
2 2 本多 男 嫌い 好き 14 10 13 B
```

```
> tbl_ht[,4]
[1] 嫌い 嫌い 好き
Levels: 嫌い 好き
```

18

```
> for (i in 6:8){print(mean(tbl_ht[,i]))}
[1] 11
[1] 7.333333
[1] 12.333333
```

```
> print("Rによるやさしい統計学")
[1] "Rによるやさしい統計学"
```

21

## 今日の自習&演習

- ◆これまでの復習
- ◆第2回レポートの作成, 質問
- ◆小テストの予想問題1, 2の検討, 質問
- ◆第9章: 9.1~9.3の例題