R

サンプルデータとスクリプト

・サンプルデータ：NICER 1.0 より学習者のエッセイデータ 287 個
・Criterion によるスコア、延べ語数、異なり語数、文数、TTR、ギローインデックス、平均単語長、平均文長、MATTR を各エッセイの特徴量として抽出
・ファイル名とともに結果をテキストファイルに保存するスクリプト

myIndices4.R

1.NICER1.0 を解凍したフォルダーの中の NICER_NNS に作業ディレクトリーを移動
2. スクリプトを実行
3. 結果を保存するファイルを作業ディレクトリーの外にファイル名をつけて「保存（作成）」
例：jpn4.txt
4.R の中に読み込む

```
> jpn4 <- read.table(choose.files())
> class(jpn4)
[1] "data.frame"
> head(jpn4)
          V1 V2  V3  V4 V5        V6       V7        V8       V9      V10
1 JPN501.txt  4 319 135 30 0.4231975 7.558549 0.5921317 4.304075 10.63333
2 JPN502.txt  4 356 161 29 0.4522472 8.532983 0.6649157 4.233146 12.27586
3 JPN503.txt  3 201 121 13 0.6019900 8.534682 0.7170149 4.746269 15.46154
4 JPN504.txt  4 260 140 27 0.5384615 8.682431 0.6877692 4.761538  9.62963
5 JPN505.txt  4 420 175 25 0.4166667 8.539126 0.6341905 3.995238 16.80000
6 JPN506.txt  3 261 124 20 0.4750958 7.675407 0.6390038 4.072797 13.05000
```

前処理

・見出しをつける

```
> names(jpn4) <- c("file", "Score", "Token", "Type", "NoS", "TTR", "GI", "MATTR", "AWL", "ASL")
> head(jpn4)
        file Score Token Type NoS       TTR       GI     MATTR      AWL      ASL
1 JPN501.txt     4   319  135  30 0.4231975 7.558549 0.5921317 4.304075 10.63333
2 JPN502.txt     4   356  161  29 0.4522472 8.532983 0.6649157 4.233146 12.27586
3 JPN503.txt     3   201  121  13 0.6019900 8.534682 0.7170149 4.746269 15.46154
4 JPN504.txt     4   260  140  27 0.5384615 8.682431 0.6877692 4.761538  9.62963
5 JPN505.txt     4   420  175  25 0.4166667 8.539126 0.6341905 3.995238 16.80000
6 JPN506.txt     3   261  124  20 0.4750958 7.675407 0.6390038 4.072797 13.05000
```

・データ数の確認

```
> nrow(jpn4)
[1] 287
```

・欠損値を調べる

```
> is.na(jpn4)
```

・欠損値がいくつかるか調べる

```
> sum(is.na(jpn4))
[1] 2
```

・欠損値を除いたデータにする

```
> jpn4.b <- na.omit(jpn4)
[1] 285
```

・jpn4.b について、各カラムを単位に操作する

```
> attach(jpn4.b)
```

こうしておけば、いちいち jpn4.b$Score としなくても Score だけでよい。

サンプルデータ（各指標）

R に読み込む

```
jpn4 <- read.csv("jpn4.csv")
```

・この状態では、欠損値 (NA) が含まれているので注意
・上記の「前処理」をしてください。