

## データリテラシーIII 最終レポートについて

自身が設定した問題意識・目的について、あるデータセットに対して、A群(分散分析、(重)回帰分析)、B群(主成分分析、因子分析、クラスター分析)のそれぞれから1つ以上の分析手法を使用して検討してください。

検討結果については、論文と同じように、タイトル・問題と目的・方法・結果と考察・引用文献の順に報告してください。必要字数等、詳しい要件は、まとめ用のフォーマットに記載していますのでこちらにしたがってください。

分析に使用した R のスクリプトファイルは必ずいっしょに提出してください(評価の対象です)。また、分析にあたって必要な関数や引数のオプションがある場合は、授業で扱っていないものについても、ぜひ自主的に調べて利用してください。その場合、努力・工夫した点として、フォーマットに付随するアンケートに、調べて利用した関数名や引数のオプションを明記した上で記入しておいてください。積極的に加点対象としたいと思います。

データセットは、下記のいずれかから選択します。

- ・自身で作成した仮想データセット
- ・自身で収集したデータセット(オープンデータ含む)
- ・立教大学からお借りしたデータセット

立教大学からお借りしたデータセットは、下記の2種類があります。

※注意:このデータセットはデータリテラシーIII の授業のために、立教大学からお借りしたものです。利用規約の関係で、授業外の目的で利用することはできませんので、取り扱いには十分注意してください。また、csv ファイルは直接ダウンロードせずに、下記コマンドにより、R に直接読み込むようにしてください。やむを得ない理由で csv ファイルをダウンロードした場合は、最終レポート提出後、速やかに削除してください。

### (1)大学生のくらしと仕事意識についてのおうかがい

- 調査の詳細は以下で確認してください。  
([調査概要](#)、[調査票](#)、[単純集計表](#))
- 下記コマンドでこの調査のデータを読み込むことができます。

```
dat <- read.csv("https://waseda.box.com/shared/static/krfyihps2mlgztxdduloz8ttaeyxoxk1.csv",  
header=T, row.names=1, na.strings=".")
```

## (2) 大学生のジェンダーと子育て意識・行動に関する調査

- 調査の詳細は以下で確認してください。

([調査概要](#)、[調査票](#)、[単純集計表](#))

- 下記コマンドでこの調査のデータを読み込むことができます。

```
dat <- read.csv("https://waseda.box.com/shared/static/yfr9jd2xd0z1zo17azbqponihuc8jun3.csv",
header=T, row.names=1, na.strings=".")
```

### \* (1)(2)の両方に共通する注意事項

- すべての設問を分析対象とする必要はありません。自身の設定した問題意識・目的に応じた項目を抜粋して検討してください。
- コーディングルールは(どの選択肢にどの数値をあてて入力されているかのルールは)、上記の調査票・単純集計表と、読み込まれる `dat` の情報を照らし合わせて確認してください。`table()`を使えば、単純集計表との対応関係がわかります。例えば、(1)のデータセットについて、`table(dat$q1a)`とすれば、

```
> table(dat$q1a)

 1  2  3  4
30 67 120 33
```

と出力されます。これを単純集計表と見比べると、1 は「1年生」、2 は「2 年生」…を表していることがわかります。

- なお、単純集計表では、欠損値と非該当が区別されていますが、演習の負担軽減のために、両者は区別せずにまとめて欠損値「.」で表現したデータセットが上記コマンドで読み込まれるようにしています。
- 欠損値は、とくに何も指定しなければ、分析ごとに、当該分析に使用する変数に欠損値がある参加者のみが分析対象から除去されます(対単位除去(`pairwise deletion`)といいます)。
- 特定のケースに絞って分析を進めたい場合は、`subset()`を利用しましょう。例えば、(1)のデータセットについて、1年生だけに絞って分析を進めたい場合は、次のようにすれば抽出できます。

```
dat1 <- subset(dat, dat$q1a==1)
```