

2020年6月

COVID-19 のデータセットから R を用いてヒストグラムを作成する【第 10 回生物統計学】

1. はじめに

今回は、前回取得した東京都の COVID-19 に関するデータセットと R を用いて、ヒストグラムを作成していこうと思います。データセットの取得に関しては、前回の「COVID-19 のデータセットを取得する【第 9 回生物統計学】(<https://www.orthomedico.jp/pdf/20200525.pdf>)」を参照してください。

2. R を用いたヒストグラムの作成

これから R を用いて、ヒストグラムを作成していこうと思います。
(R のセットアップ方法をいろいろなサイトにあります。)

R でデータを読み込む前に、データセットの下準備をしてください。手順は以下の通りです。

【データセットの下準備】

- ①取得した csv ファイルをメモ帳で開く。
- ②新規のメモ帳を開き、そこに①のメモ帳のデータを貼り付ける。
- ③②で作成したメモ帳に任意のファイル名を付けて保存する。

東京都の HP から取得したファイルを直接用いたかったのですが、うまく動きませんでした(´ω´)
上記の方法ならうまくデータセットを取り込めますよ(*^▽^)/

データセットの下準備ができれば、下記に示した R のスクリプトでヒストグラムが作成できます。

```
#作業ディレクトリを変更する
setwd("<ディレクトリに設定するアドレス>")

#現在の作業ディレクトリを確認する
getwd()

#データセットを読み込む
x<-read.table("130001_tokyo_covid19_patients.txt", header=TRUE, sep=",",
na.strings=c(" "), stringsAsFactors = FALSE, fileEncoding="utf-8")

#5 行目までのデータを確認してみる
head(x, 5)
```



```
#クラス確認
```

```
sapply(x, class)
```

```
#lubridate をインストールする
```

```
install.packages("lubridate")
```

```
install.packages("Rcpp")
```

```
#lubridate を展開する
```

```
library("lubridate")
```

```
#日毎の感染者数のヒストグラムを作成する場合
```

```
hist(date(x$公表_年月日),breaks="days", freq=T, main="新規患者に関する報告件数の推移",  
xlab="公表月日", ylab="件数", ylim=c(0,250))
```

```
#月毎の感染者数のヒストグラムを作成する場合
```

```
hist(date(x$公表_年月日),breaks="month", freq=T, main="新規患者に関する報告件数の推移",  
xlab="公表月日", ylab="件数", ylim=c(0,4000))
```

結果を出力できましたか?

色や大きさも設定できるので、綺麗な図作りを試してみるのも良いですね。

R では、データセットから様々なグラフが作成できます。COVID-19 のデータ以外にも都や国が収集したデータや実際に臨床試験で取得したデータに対し R 用いて、表を作成してみたいかかでしょうか。販促資料として利用できる魅力的なグラフが簡単に作成できるかもしれませんね。

3. まとめ

今回は COVID-19 のデータセットを調査してみました。また、次回も生物統計に関する情報を紹介しますのでお楽しみに(^^#)

4. 動作環境確認

- R x64 3.5.3
- R x64 3.6.1