FOG time to events analysis

Christophe Mpaga

Contents

Executive summary

In this project, we will be using data from kaggle.We aim to analysis time to freezing of gait (FOG). FOG is a pattern occurring in patient with Parkinson diseases. It indicates kinetic inability and impairment during gait for instance. Some indicative events like walking hesitation, turning body could be observed and help to detect FOG occurence.

let's join all metadata tables before diving into analysis.

- defog_metadata.csv Identifies each series in the tdcsfog dataset by a unique Subject, Visit, Test, Medication condition.
  - `Visit` Lab visits consist of a baseline assessment, two post-treatment assessments for different treatment stages, and one follow-up assessment.
  - `Test` Which of three test types was performed, with 3 the most challenging.
  - `Medication` Subjects may have been either off or on anti-parkinsonian medication during the recording.
- subjects.csv Metadata for each Subject in the study, including their `Age` and `Sex` as well as:
  - `Visit` Only available for subjects in the daily and defog datasets.
  - `YearsSinceDx` Years since Parkinson's diagnosis.
  - `UPDRSIIIOn/UPDRSIIIOff` Unified Parkinson's Disease Rating Scale score during on/off medication respectively. `NFOGQ` Self-report FoG questionnaire score. See: https://pubmed.ncbi.nlm.nih.gov/19660949/
- events.csv Metadata for each FoG event in all data series. The event times agree with the labels in the data series.
  - `Id` The data series the event occured in.
  - `Init` Time (s) the event began.

  - `Completion` Time (s) the event ended.
  - `Type` Whether StartHesitation, Turn, or Walking.
  - `Kinetic` Whether the event was kinetic (1) and involved movement, or akinetic (0) and static.

- tasks.csv Task metadata for series in the defog dataset. (Not relevant for the series in the tdcsfog or daily datasets.)

  - `Id` The data series where the task was measured.
  - `Begin` Time (s) the task began.
  - `End` Time (s) the task ended.
    Task One of seven tasks types in the DeFOG protocol, described on this page.

Note: we are going to consider only defog condition subjects. # data inspection and preprocessing
data structure

```
## [1] 80509    18
```

```
## # A tibble: 3 x 18
##   Id        Init Compl~1 Type  Kinetic Begin   End Task  Subject Visit.x Medic~2
##   <chr>    <dbl>   <dbl> <chr>   <dbl> <dbl> <dbl> <chr> <chr>     <dbl> <chr>
## 1 02ab235~ 1929.   1933. <NA>       NA   10   190. Rest1 e1f62e        2 on
## 2 02ab235~ 1929.   1933. <NA>       NA   10   190. Rest1 e1f62e        2 on
## 3 02ab235~ 1929.   1933. <NA>       NA  211.  272. Rest2 e1f62e        2 on
## # ... with 7 more variables: Visit.y <dbl>, Age <dbl>, Sex <chr>,
## #   YearsSinceDx <dbl>, UPDRSIII_On <dbl>, UPDRSIII_Off <dbl>, NFOGQ <dbl>, and
## #   abbreviated variable names 1: Completion, 2: Medication
```

Here we will refer to Kinetic as status (will change it later) and will compute duration in following cell.

Adding duration column : Completion - Init

```
## # A tibble: 3 x 16
##   Id     Type  Kinetic Task  Subject Visit.x Medic~1 Visit.y   Age Sex   Years~2
##   <chr>  <chr>   <dbl> <chr> <chr>     <dbl> <chr>     <dbl> <dbl> <chr>   <dbl>
## 1 02ab2~ <NA>       NA Rest1 e1f62e        2 on            1    79 F           8
## 2 02ab2~ <NA>       NA Rest1 e1f62e        2 on            2    79 F           8
## 3 02ab2~ <NA>       NA Rest2 e1f62e        2 on            1    79 F           8
## # ... with 5 more variables: UPDRSIII_On <dbl>, UPDRSIII_Off <dbl>,
## #   NFOGQ <dbl>, eventsDuration <dbl>, tasksDuration <dbl>, and abbreviated
## #   variable names 1: Medication, 2: YearsSinceDx
```

Data analysis : 1D EDA

How many unique subjects are in this dataset ?

```
## [1] 124
```

There are 124 distinct subjects in this study.

What is the mean duration of events and tasks ?

```
## events mean duration : 4.84926 s
```

```
## tasks mean duration : 45.76702 s
```

```
## # A tibble: 3 x 16
##   Id     Type  Kinetic Task  Subject Visit.x Medic~1 Visit.y   Age Sex   Years~2
##   <chr>  <chr>   <dbl> <chr> <chr>     <dbl> <chr>     <dbl> <dbl> <chr>   <dbl>
## 1 f9fc6~ Turn        1 Hots~ 040587        1 on            1    75 M          26
## 2 f9fc6~ Turn        1 Hots~ 040587        1 on            2    75 M          26
## 3 f9fc6~ Turn        1 Hots~ 040587        1 on            1    75 M          26
## # ... with 5 more variables: UPDRSIII_On <dbl>, UPDRSIII_Off <dbl>,
## #   NFOGQ <dbl>, eventsDuration <dbl>, tasksDuration <dbl>, and abbreviated
## #   variable names 1: Medication, 2: YearsSinceDx
```

summary stats

```
##       Id                Type              Kinetic          Task          
##  Length:80509       Length:80509       Min.   :0.00    Length:80509      
##  Class :character   Class :character   1st Qu.:0.00    Class :character  
##  Mode  :character   Mode  :character   Median :1.00    Mode  :character  
##                                        Mean   :0.67                      
##                                        3rd Qu.:1.00                      
##                                        Max.   :1.00                      
##                                        NA's   :42498                     
##    Subject             Visit.x       Medication          Visit.y      
##  Length:80509       Min.   :1.000   Length:80509       Min.   :1.000  
##  Class :character   1st Qu.:1.000   Class :character   1st Qu.:1.000  
##  Mode  :character   Median :2.000   Mode  :character   Median :2.000  
##                     Mean   :1.683                      Mean   :1.502  
##                     3rd Qu.:2.000                      3rd Qu.:2.000  
##                     Max.   :2.000                      Max.   :2.000  
##                                                                       
##       Age            Sex             YearsSinceDx    UPDRSIII_On    
##  Min.   :28.00   Length:80509       Min.   : 1.50   Min.   :13.00  
##  1st Qu.:61.00   Class :character   1st Qu.: 7.00   1st Qu.:34.00  
##  Median :69.00   Mode  :character   Median :13.00   Median :38.00  
##  Mean   :67.54                      Mean   :12.55   Mean   :38.65  
##  3rd Qu.:73.00                      3rd Qu.:16.00   3rd Qu.:47.00  
##  Max.   :82.00                      Max.   :30.00   Max.   :57.00  
##                                                                    
##   UPDRSIII_Off       NFOGQ       eventsDuration    tasksDuration     
##  Min.   :18.00   Min.   :10.00   Min.   :  0.119   Min.   :  0.003  
##  1st Qu.:42.00   1st Qu.:19.00   1st Qu.:  0.971   1st Qu.: 13.230  
##  Median :47.00   Median :22.00   Median :  2.243   Median : 23.520  
##  Mean   :46.44   Mean   :21.77   Mean   :  4.849   Mean   : 45.767  
##  3rd Qu.:52.00   3rd Qu.:24.00   3rd Qu.:  5.377   3rd Qu.: 54.440  
##  Max.   :76.00   Max.   :28.00   Max.   :144.565   Max.   :356.720  
##
```

Kinetic has 0.53 % of NA.

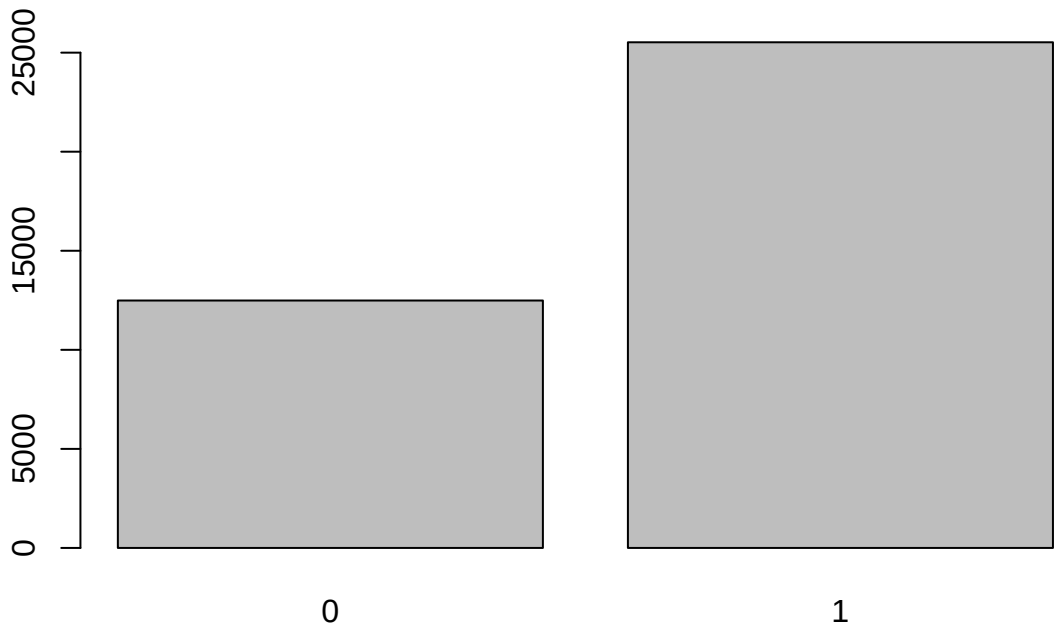
Kinetic

How many trials has missing kinetic/status ?

```
## [1] 0.53
```

53% of trials has missing Kinetic(status).

Among remaining 47% of trials, how often does a kinetic event occurs ?
Kinetic graphical summary



Kinetic numerical summary

```
## Kinetic
##    0    1
## 0.33 0.67
```

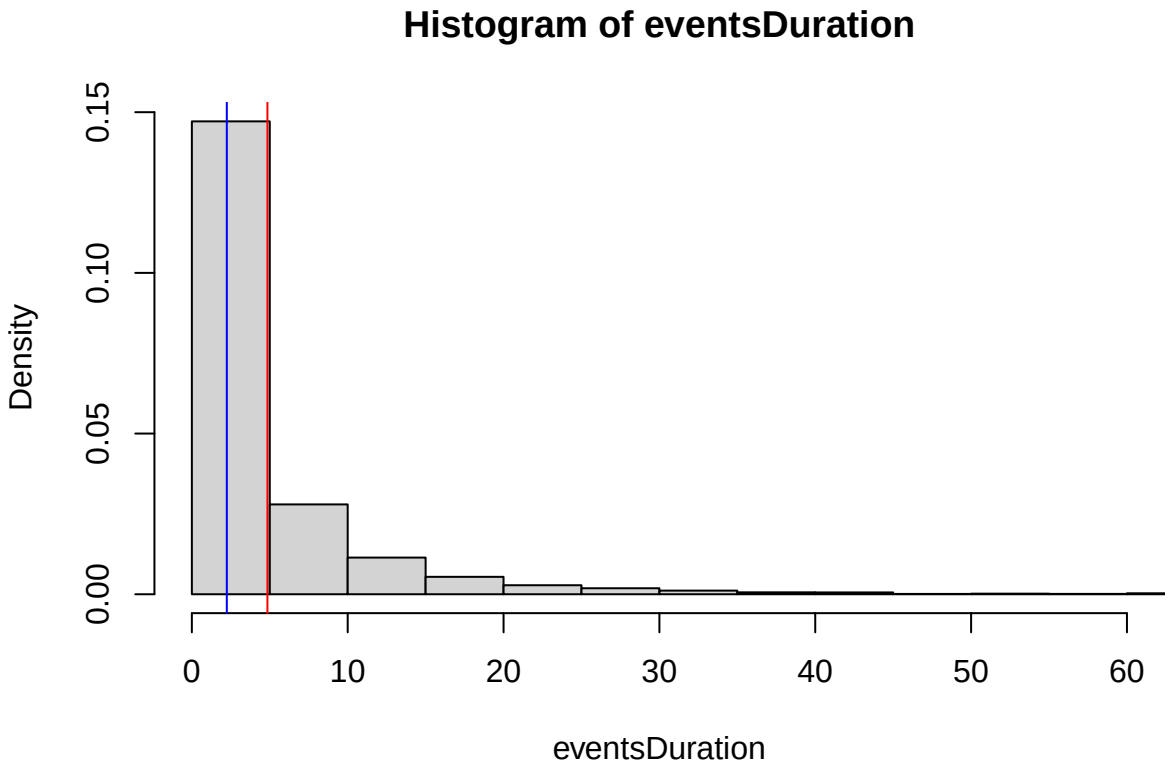
0.33 % of events has been censored.

Note : Given that each event is indicative of FOG, we will count each of them equally for the moment.

events type numerical summary

```
## Type
## StartHesitation            Turn         Walking
##            0.01            0.81            0.18
```

The most frequent events is Turn. (81% of the observed events).
fog Duration graphical summary.



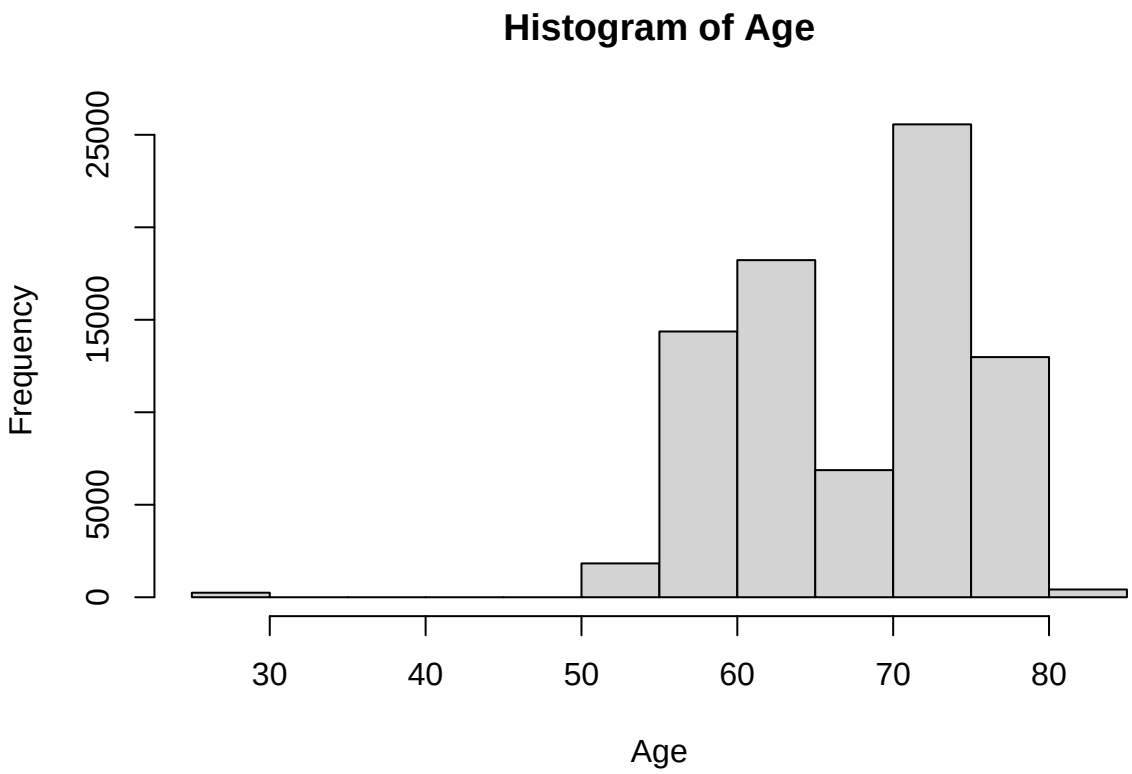
Events duration is asymetric and right skewed.
EventsDuration Numerical summary

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.119   0.971   2.243   4.849   5.377 144.565
```

At least 50% of cases has `2.243`s events duration, and events duration ranges between 0 .11 and 581 seconds.

Age

What is the median age ?
Age graphical summary.



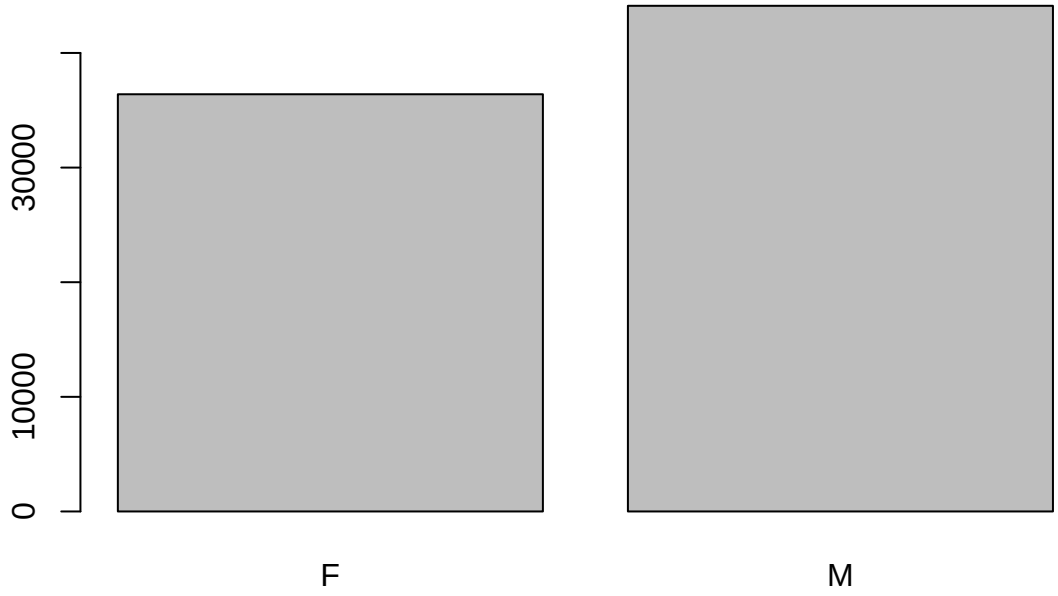
Age is not normally distributed, bimodal with some outliers.
Age numerical summary.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   28.00   61.00   69.00   67.54   73.00   82.00
```

At least 50% of subjects are 69 years old.

Sex

sex graphical summary.



sex numerical summary

```
## Sex
##    F    M
## 0.45 0.55
```

There is almost 0.55% of men in this cohort.

tasks

to be continued !!!