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 disease. 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.

Introduction

In this application we aim to analyse time to event in Parkinson patients.Our main goal is to analyse and explain the influence/importance of some features in the occurrence of freezing of gait(FOG). To achieve this, we model our data under kaplan-Meier model assumption.

Data preparation, inspection and preprocessing.

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.

Remove **Visit** from **subjects** table , as we do not use it

Note: we are going to consider only defog condition subjects.

Data structure

```
## data table dimension
```

```
## [1] 80509    17
```

```
##
##  first 3 rows
```

```
## # A tibble: 3 x 17
##   Id           Init Compl~1 Type  Kinetic Begin   End Task  Subject Visit Medic~2
##   <chr>       <dbl>   <dbl> <chr>   <dbl> <dbl> <dbl> <chr> <chr>   <dbl> <chr>
## 1 02ab235146 1929.    1933. <NA>       NA    10  190. Rest1 e1f62e      2 on
## 2 02ab235146 1929.    1933. <NA>       NA    10  190. Rest1 e1f62e      2 on
## 3 02ab235146 1929.    1933. <NA>       NA   211.  272. Rest2 e1f62e      2 on
## # ... with 6 more variables: Age <dbl>, Sex <chr>, YearsSinceDx <dbl>,
## #   UPDRSIII_On <dbl>, UPDRSIII_Off <dbl>, NFOGQ <dbl>, and abbreviated
## #   variable names 1: Completion, 2: Medication


##
##   last 3 rows


## # A tibble: 3 x 17
##   Id           Init Compl~1 Type  Kinetic Begin   End Task  Subject Visit Medic~2
##   <chr>       <dbl>   <dbl> <chr>   <dbl> <dbl> <dbl> <chr> <chr>   <dbl> <chr>
## 1 f9fc61ce85 1173.    1173. Turn        1 1087. 1105. Hots~ 040587      1 on
## 2 f9fc61ce85 1173.    1173. Turn        1 1152. 1190. Hots~ 040587      1 on
## 3 f9fc61ce85 1173.    1173. Turn        1 1152. 1190. Hots~ 040587      1 on
## # ... with 6 more variables: Age <dbl>, Sex <chr>, YearsSinceDx <dbl>,
## #   UPDRSIII_On <dbl>, UPDRSIII_Off <dbl>, NFOGQ <dbl>, and abbreviated
## #   variable names 1: Completion, 2: Medication


## # A tibble: 3 x 17
##   Id           Init Compl~1 Type  Kinetic Begin   End Task  Subject Visit Medic~2
##   <chr>       <dbl>   <dbl> <chr>   <dbl> <dbl> <dbl> <chr> <chr>   <dbl> <chr>
## 1 f9fc61ce85 1173.    1173. Turn        1 1087. 1105. Hots~ 040587      1 on
## 2 f9fc61ce85 1173.    1173. Turn        1 1152. 1190. Hots~ 040587      1 on
## 3 f9fc61ce85 1173.    1173. Turn        1 1152. 1190. Hots~ 040587      1 on
## # ... with 6 more variables: 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.

Ad event and tasks duration columns :

- `eventsDuration <- Completion - Init`
- `tasksDuration <- Begin - end`

Data analysis : 1D EDA

Subjects

How many unique subjects are in this dataset ?

There are 44 distinct subjects in this study.

Duration

What is the mean duration of events and tasks ?

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

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

```
## spc_tbl_ [80,509 x 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ Id           : chr [1:80509] "02ab235146" "02ab235146" "02ab235146" "02ab235146" ...
## $ Type         : chr [1:80509] NA NA NA NA ...
## $ Kinetic      : num [1:80509] NA NA NA NA NA NA NA NA NA NA ...
## $ Task         : chr [1:80509] "Rest1" "Rest1" "Rest2" "Rest2" ...
## $ Subject      : chr [1:80509] "e1f62e" "e1f62e" "e1f62e" "e1f62e" ...
## $ Visit        : num [1:80509] 2 2 2 2 2 2 2 2 2 2 ...
## $ Medication   : chr [1:80509] "on" "on" "on" "on" ...
## $ Age          : num [1:80509] 79 79 79 79 79 79 79 79 79 79 ...
## $ Sex          : chr [1:80509] "F" "F" "F" "F" ...
## $ YearsSinceDx : num [1:80509] 8 8 8 8 8 8 8 8 8 8 ...
## $ UPDRSIII_On  : num [1:80509] 52 38 52 38 52 38 52 38 52 38 ...
## $ UPDRSIII_Off : num [1:80509] 55 64 55 64 55 64 55 64 55 64 ...
## $ NFOGQ        : num [1:80509] 28 28 28 28 28 28 28 28 28 28 ...
## $ eventsDuration: num [1:80509] 3.36 3.36 3.36 3.36 3.36 ...
## $ tasksDuration : num [1:80509] 180.5 180.5 60.3 60.3 16.5 ...
## - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_character(),
##   ..   Init = col_double(),
##   ..   Completion = col_double(),
##   ..   Type = col_character(),
##   ..   Kinetic = col_double()
##   .. )
## - attr(*, "problems")=<externalptr>
```

Kinetic has 53 % of NA.

Kinetic / Events

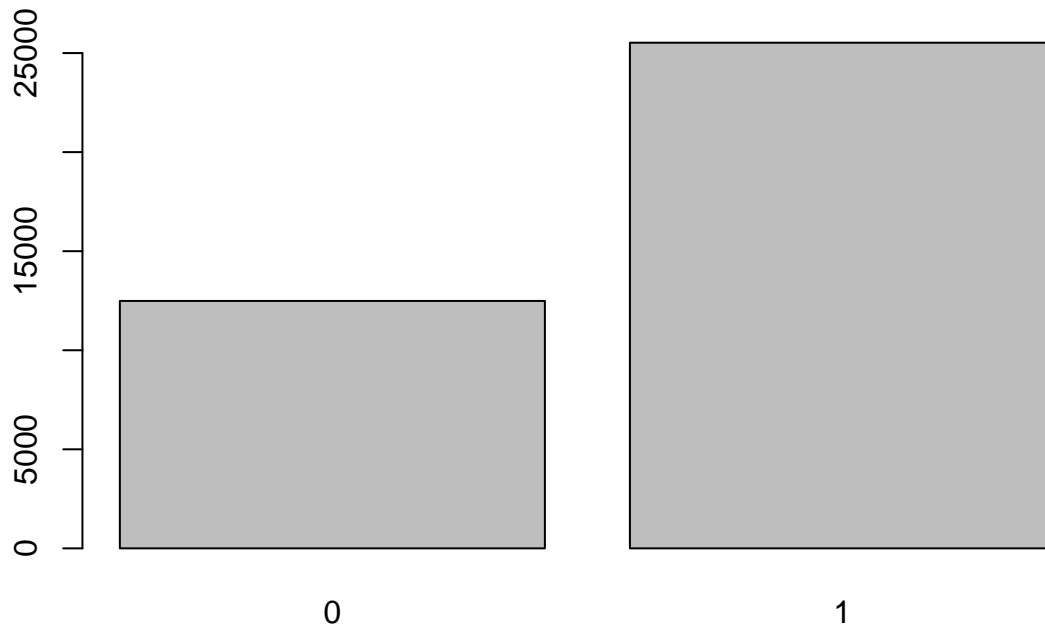How many trials has missing kinetic/status ?

53% of trials has missing Kinetic(status).

filter for missing kinetic entries/cases.

```
##
## new table dimension
```

```
## [1] 42498     15
```

4

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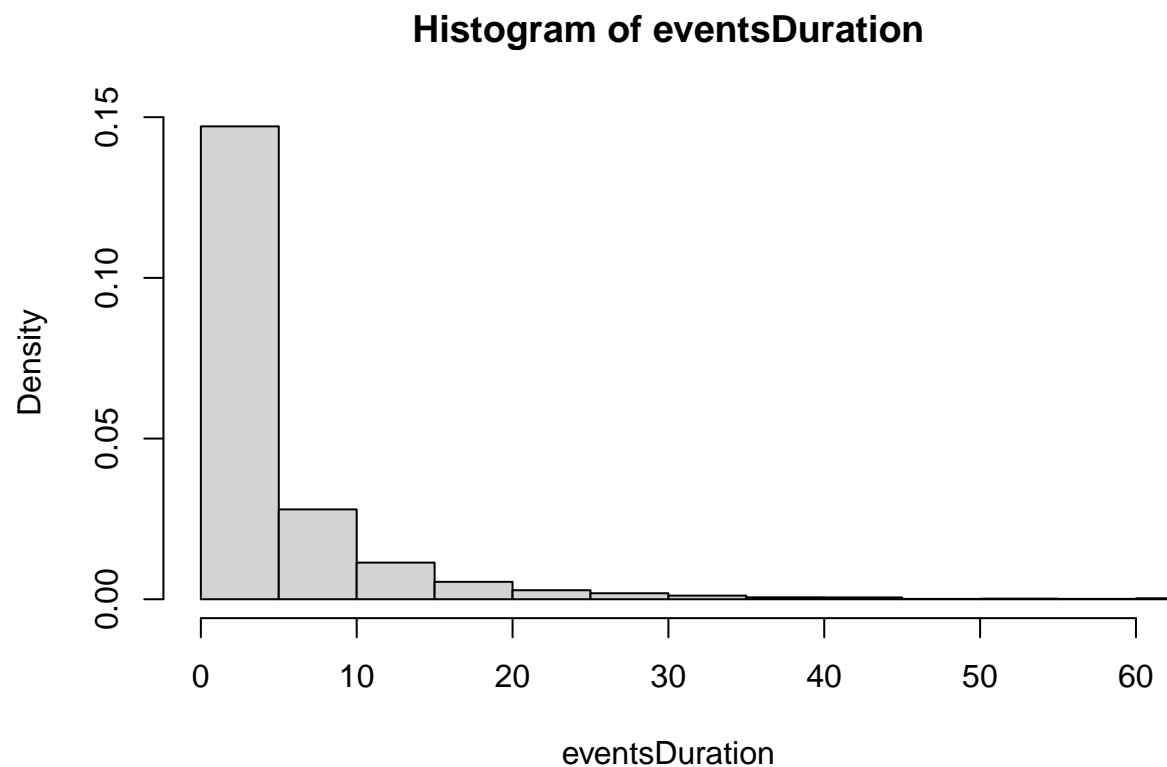Kinetic graphical summary

Kinetic numerical summary

33 % of events has been censored/not observed.

Note : Given that each event is indicative of FOG, we will just gather a nd consider them in one class (Kinetic / events).

events type numerical summary

events Duration graphical summary.

Histogram of eventsDuration

Events duration is asymetric and right skewed.
Many events duration are of high values.

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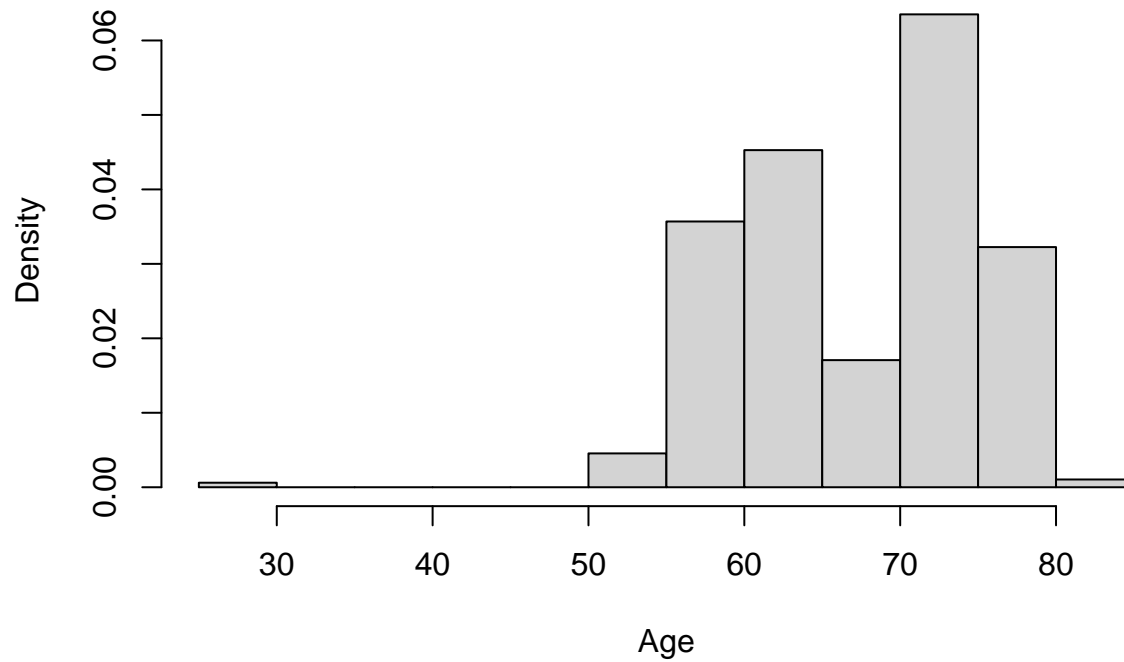
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.119   0.971   2.243   4.849   5.377 144.565
```

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

Age

What is the median age ?
Age graphical summary.
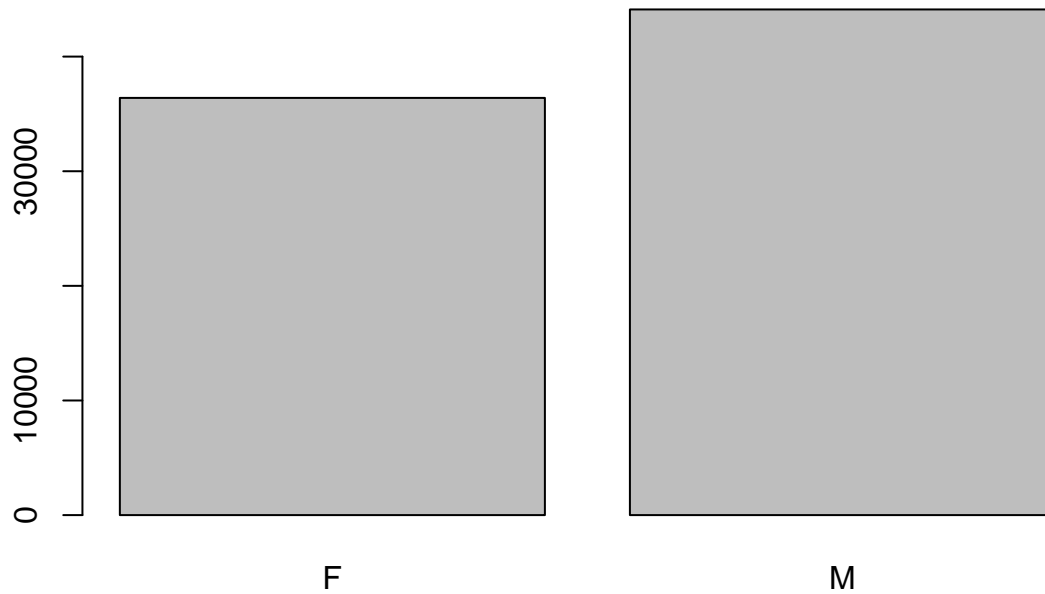
Histogram of Age

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

There is almost 54.7926319% of men in this cohort.

Tasks

What kind of tasks has been performed ?

```
##  [1] "Rest1"      "Rest2"      "4MW"        "4MW-C"      "MB1"
##  [6] "MB2a"       "MB2b"       "MB3-R"      "MB3-L"      "MB4"
## [11] "MB5"        "MB6-R"      "MB6-L"      "MB7"        "MB8"
## [16] "MB9"        "MB10"       "MB11"       "MB12"       "MB13"
## [21] "TUG-ST"     "TUG-DT"     "TUG-C"      "Turning-ST" "Turning-DT"
## [26] "Turning-C"  "Hotspot1"   "Hotspot1-C" "Hotspot2"   "Hotspot2-C"
## [31] "MB6"
```

how many tasks had been performed ?

There had been 31 task performed.

Visit

How many rounds of visit did the patient had ?

```
## [1] 2
```

What is the proportion of patient within each number of visit round ?

```
## Visit
##    1    2
## 0.32 0.68
```

32% of patients had one Visit, while the remaining had 2.

Medication

How many subjects are under parkinson-medication ?

49.72% are under medication.

To be continued !!!!