# 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. Here FOG is the event of interest and it is expected within the execution of a given task, or during daily life activity. Here we will analyse data from lab tasks. Our main goal is to analyse and explain the influence/importance of some features in the occurrence of freezing of gait(FOG) from the beginning to the end of the task. 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

Here we will refer to Kinetic as status (will change it later) and init as Time , then we will compute duration .

**Ad event and tasks duration columns features:**

- `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.

### Kinetic / Events

How many trials has missing kinetic/status ?

```
## [1] 0.48
```

48% of trials has missing Kinetic(status).

filter for missing kinetic entries/cases.

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

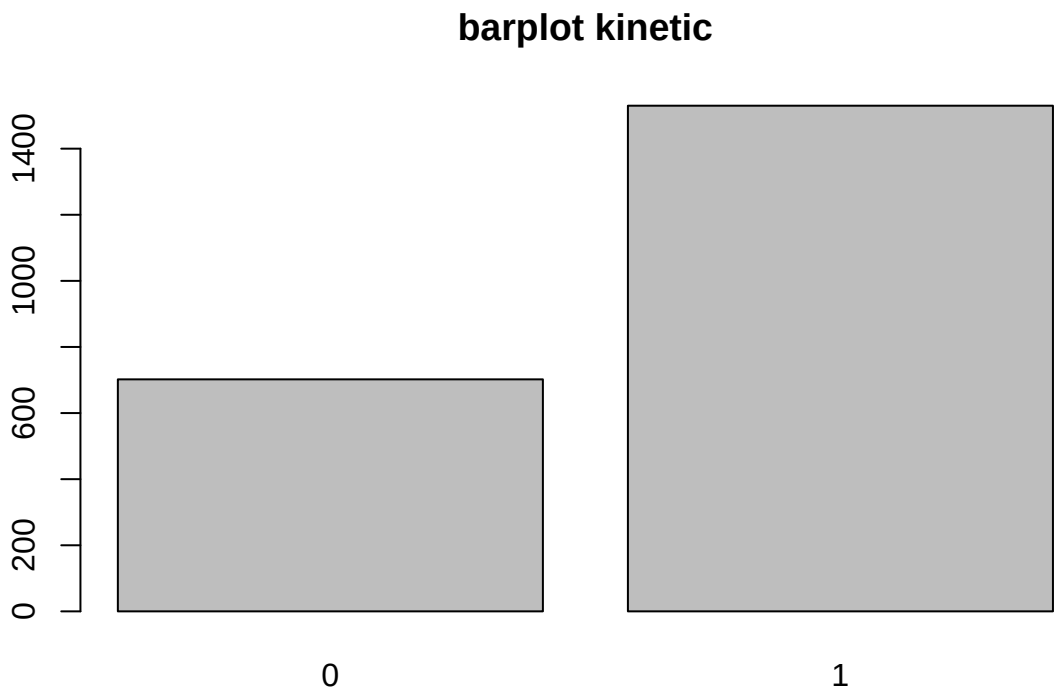
```
## [1] 2232   20
```

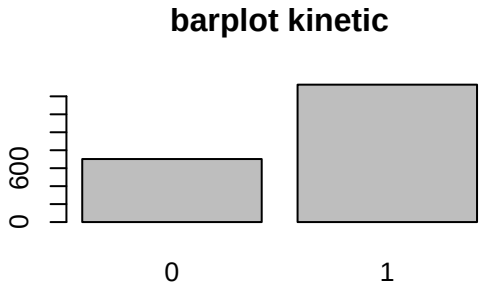
```
## # A tibble: 3 x 20
##   Id          Init Compl~1 Type  Kinetic Begin   End Task  Subject Visit Medic~2
##   <chr>      <dbl>   <dbl> <chr>   <dbl> <dbl> <dbl> <chr> <chr>   <dbl> <chr>
## 1 02ea782681 1377.   1378. Turn        1 1371. 1393. Turn~ ae2d35      2 on
## 2 02ea782681 1377.   1378. Turn        1 1371. 1393. Turn~ ae2d35      2 on
## 3 02ea782681 1466.   1467. Turn        1 1461. 1472. Hots~ ae2d35      2 on
## # ... with 9 more variables: Age <dbl>, Sex <chr>, YearsSinceDx <dbl>,
## #   UPDRSIII_On <dbl>, UPDRSIII_Off <dbl>, NFOGQ <dbl>, eventsDuration <dbl>,
## #   tasksDuration <dbl>, Time <dbl>, and abbreviated variable names
## #   1: Completion, 2: Medication
```

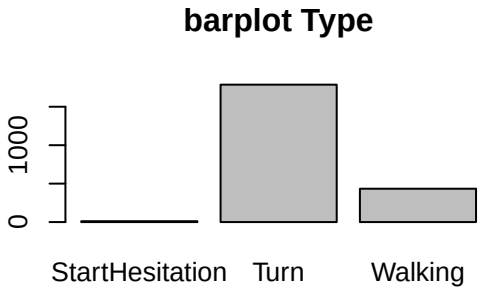
```
## [1] 2232   20
```

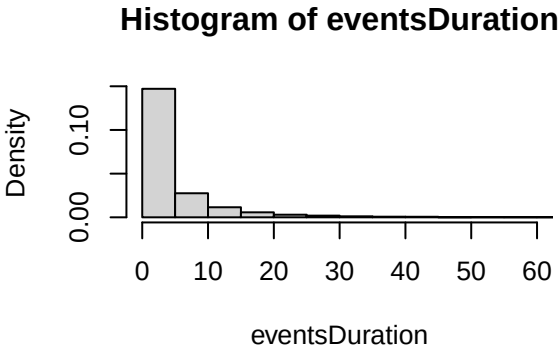
```
## [1] 38 15
```

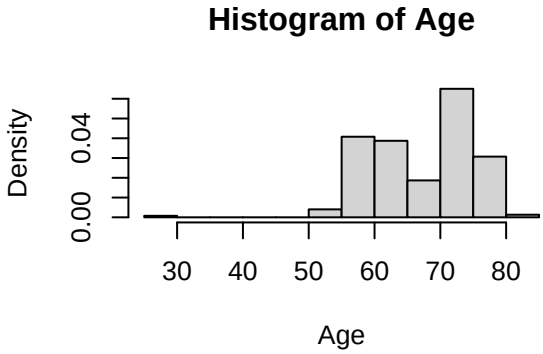
Among remaining 52% of trials. What is the proportion of observed Kinetic events ?
Kinetic graphical summary






multiple plot

Kinetic numerical summary

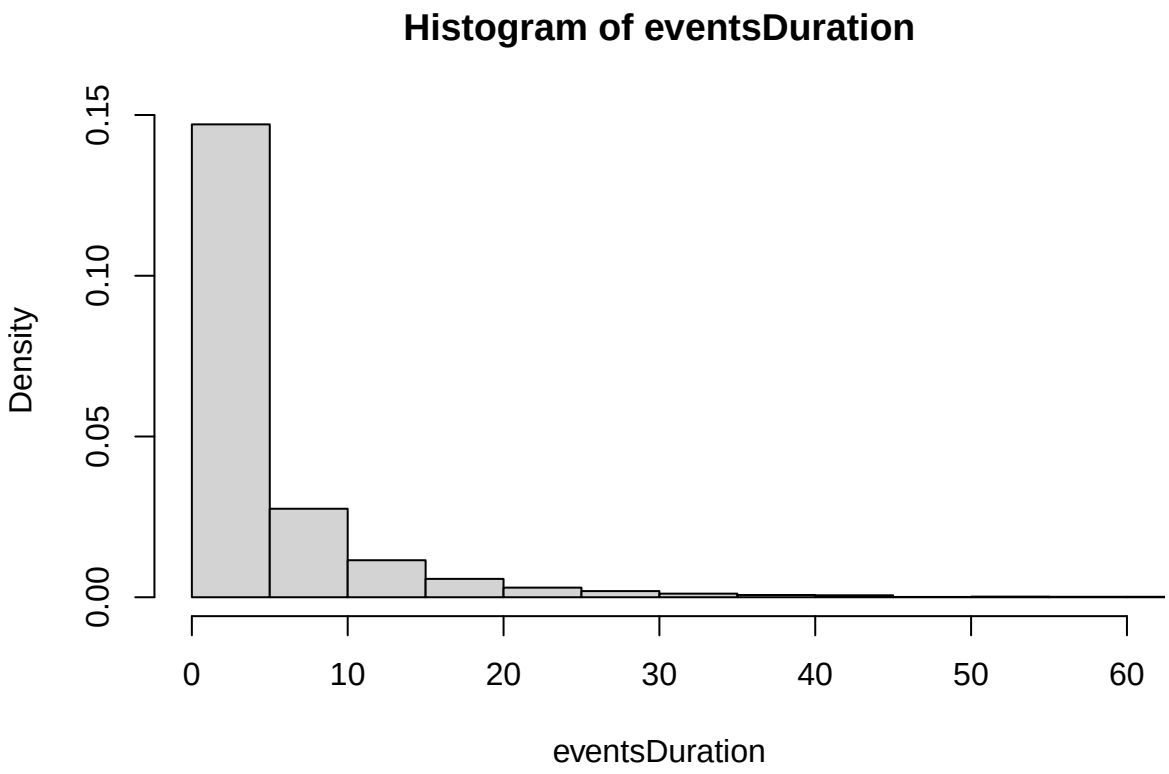
69 % of events has been observed.

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


events type numerical summary

```
## Type
## StartHesitation            Turn         Walking
##            0.00            0.80            0.19
```

events Duration graphical summary.




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.
##   0.1190   0.9732   2.2300   4.7992   5.4370 144.5650
```

At least 50% of cases has 2.23s events duration, and events duration ranges between 0 .11 and 581 seconds.

## Duration

What is the mean duration of events, tasks and Time ?

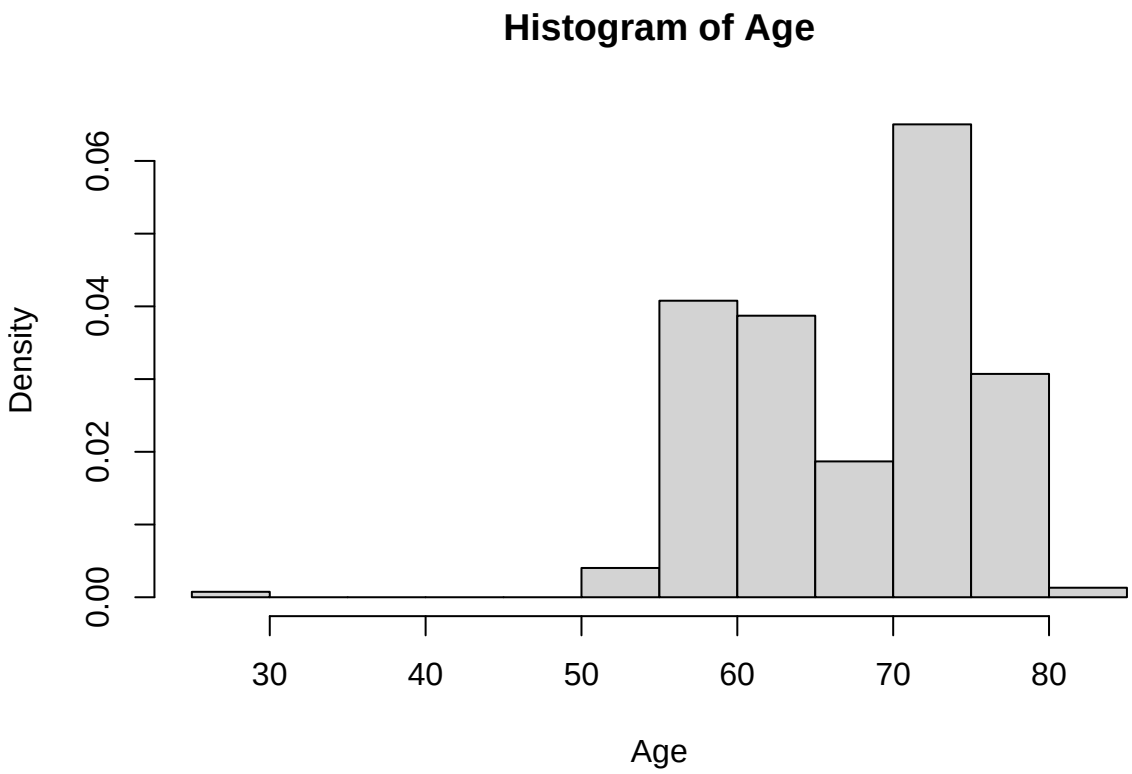
```
## events mean duration : 4.799225 s
```

```
## tasks mean duration : 101.1519 s
```

```
## mean time to events duration : 0.003139747 s
```

## 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   67.44   73.00   82.00
```

At least half of subjects are 69 years old.

### Sex

Sex graphical summary.

Sex numerical summary

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

There is almost 55.1498127% of men in this cohort.

### Tasks

What kind of tasks has been performed ?

```
##  [1] "TUG-DT"     "TUG-C"      "Turning-DT" "Hotspot2"   "Hotspot2-C"
##  [6] "Hotspot1"   "Turning-ST" "4MW"        "4MW-C"      "TUG-ST"
## [11] "Turning-C"  "Hotspot1-C" "MB10"       "MB11"       "MB13"
## [16] "MB12"
```

how many tasks had been performed ?

There had been 16 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
## 0.37 0.63
```

37% of patients had one Visit, while the remaining had 2.

## Medication

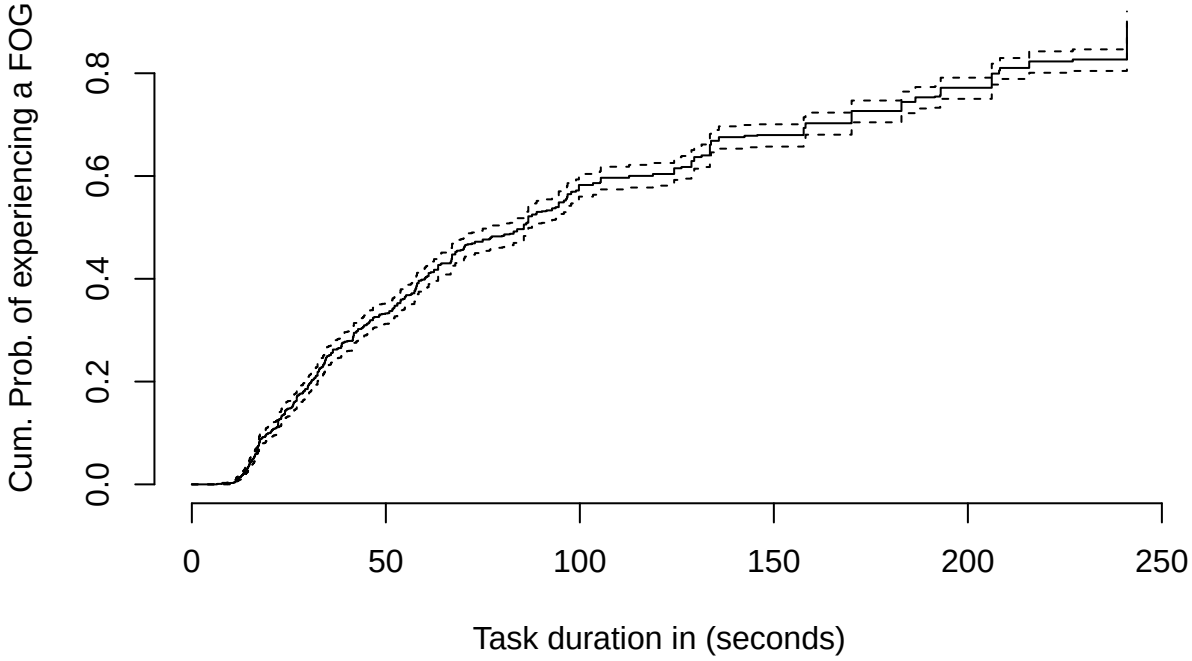
How many subjects are under parkinson-medication ?

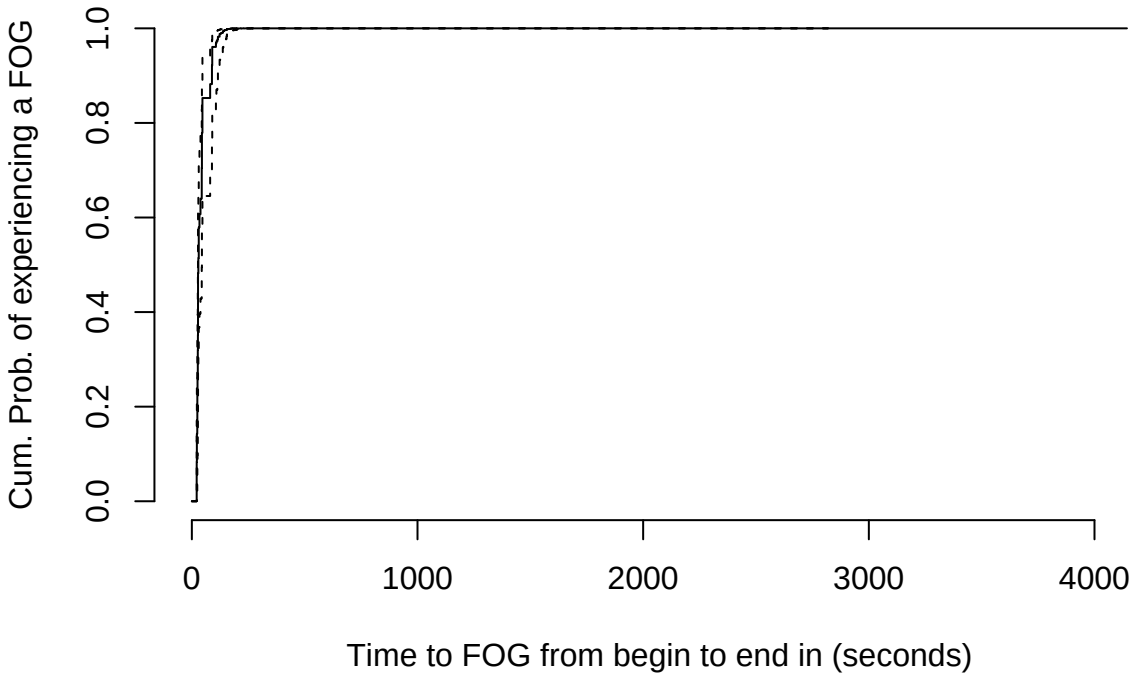
```
## Medication
##   off    on
## 68.89 31.11
```

31.11% are under medication. ##

**2D EDA**

## Modelling and Analysis : overall time to FOG.






# summary fit

```
## Call: survfit(formula = Surv(tasksDuration, Kinetic) ~ 1, data = fog)
##
##         n events median 0.95LCL 0.95UCL
## [1,] 2232   1530   85.6      77    87.7

## Call: survfit(formula = Surv(Begin, End, Kinetic) ~ 1, data = fog)
##
##      records n.max n.start events median 0.95LCL 0.95UCL
## [1,]    2232   190      36   1530   28.9    27.4    44.7
```

Half of patients experiment FOG within 85.6 s.

### probability of FOG by 2 minutes.

```
## Call: survfit(formula = Surv(Begin, End, Kinetic) ~ 1, data = fog)
##
##  time n.risk n.event censored survival  std.err lower 95% CI upper 95% CI
##   300     29     122       15 1.56e-06 2.18e-06     1.01e-07     2.41e-05

## Call:  coxph(formula = Surv(Begin, End, Kinetic) ~ 1, data = fog)
##
## Null model
##   log likelihood= -6413.235
##   n= 2232
```

## Discussion and Conclusion

## References

**To be continued !!!!**