Portada Bien Fachera

Contextualización, objetivos y descripción de los datos

La enfermedad de Parkinson es un transtorno neurodegenerativo crónico caracterizado por los temblores, rigidez y disminución de la mobilidad. Esta enfermedad se debe a un déficit en la secreción de dopamina, hormona liberada por las terminaciones nerviosas de la sustancia negra. A veces comienza con un temblor apenas perceptible en una sola mano. En las etapas iniciales de la enfermedad de Parkinson, el rostro puede tener una expresión leve o nula. Es posible que los brazos no se balanceen al caminar. El habla puede volverse suave o incomprensible. Los síntomas de la enfermedad de Parkinson se agravan a medida que la enfermedad progresa con el tiempo.

A pesar de que la enfermedad de Parkinson no tiene cura, los medicamentos pueden reducir o atenuar notablemente los síntomas. En ocasiones, el médico puede sugerir realizar una cirugía para regular determinadas zonas del cerebro.

Esta enfermedad representa el segundo trastorno neurodegenerativo por su frecuencia, sólo por detrás del Alzheimer. Está extendida por todo el mundo y puede desarrollarse en ambos sexos, afectando a entre un 1% a un 2% de la población mayor de 60 años.

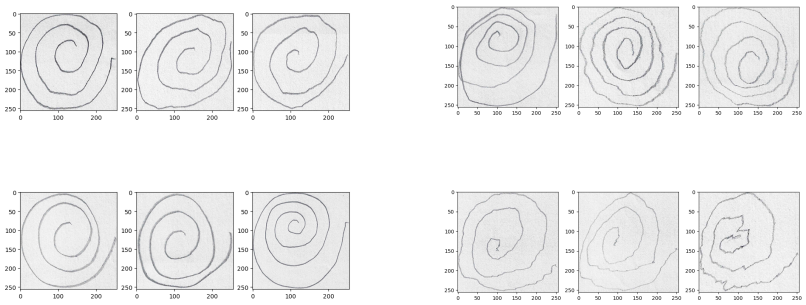

Figure 1: Test de la espiral en pacientes sanos (izquierda) y pacientes con Parkinson (derecha)

Hemos obtenido el dataset con el que vamos a trabajar de la web Kaggle, bajo el título “Early Biomarkers of Parkinson’s Disease”, que podría traducirse como “Biomarcadores prematuros de la enfermedad de Parkinson” Los datos provienen originalmente de un paper publicado a principios de 2017 en nature.com, que puede consultarse en el enlace https://www.nature.com/articles/s41598-017-00047-5.pdf

Hemos elegido este tema porque nos parece que la investigación científica y médica es una área que se beneficia mucho del análisis de datos ya que abunda la información puesto que se realizan todo tipo de pruebas y diagnósticos a los pacientes. Además, estudios de este tipo están enfocados en mejorar la calidad de vida de los pacientes afectados por la enfermedad, y claramente nos parece que merece la pena trabajar de cara a éste objetivo.

El conjunto de datos incluye 30 pacientes con enfermedad de Parkinson (EP) temprana no tratada, 50 pacientes con trastorno de conducta del sueño REM (RBD), que tienen un alto riesgo de desarrollar la enfermedad de Parkinson, y 50 controles sanos (HC).

Todos los pacientes fueron evaluados clínicamente por un neurólogo profesional con experiencia en trastornos del movimiento. Todos los sujetos fueron examinados durante una sola sesión con un especialista del habla. Éstos realizaron la lectura de un texto estandarizado, fonéticamente equilibrado de 80 palabras y monólogos sobre sus intereses, trabajo, familia o actividades actuales durante aproximadamente 90 segundos. Las características del habla fueron analizadas automáticamente por Jan Hlavnička et al.

Con el análisis de éste conjunto de datos se pretende:

- Hallar biomarcadores de la enfermedad de Parkinson en los distintos pacientes estudiados
- Clasificar a los pacientes en grupos de riesgo
- Distinguir qué rasgos están más estrechamente relacionados con el desarrollo del Parkinson

Descripción de los datos

```
data <- read.csv("dataset.csv")
data <- as_tibble(data)
str(data)
```

```
## tibble [130 x 65] (S3: tbl_df/tbl/data.frame)
##  $ Participant..code                                          : chr [1:130] "PD01" "PD02" "PD03" "
##  $ Age...years.                                               : int [1:130] 58 68 68 75 61 58 79 5
##  $ Gender                                                     : chr [1:130] "F" "F" "M" "M" ...
##  $ Positive..history..of..Parkinson..disease..in..family      : chr [1:130] "No" "No" "No" "No" ..
##  $ Age..of..disease..onset...years.                           : chr [1:130] "56" "67" "67" "73" ..
##  $ Duration..of..disease..from..first..symptoms...years.      : chr [1:130] "2" "1" "1" "2" ...
##  $ Antidepressant..therapy                                    : chr [1:130] "No" "No" "No" "No" ..
##  $ Antiparkinsonian..medication                               : chr [1:130] "No" "No" "No" "No" ..
##  $ Antipsychotic..medication                                  : chr [1:130] "No" "No" "No" "No" ..
##  $ Benzodiazepine..medication                                 : chr [1:130] "No" "No" "No" "No" ..
##  $ Levodopa..equivalent...mg.day.                             : int [1:130] 0 0 0 0 0 0 0 0 0 0 ..
##  $ Clonazepam...mg.day.                                       : num [1:130] 0 0 0 0 0 0 0 0 0 0 ..
##  $ Overview..of..motor..examination...Hoehn.....Yahr..scale.....: chr [1:130] "1.5" "2.5" "2" "2" ..
##  $ Overview..of..motor..examination...UPDRS..III..total.....  : chr [1:130] "8" "22" "19" "24" ...
##  $ X18...Speech                                               : chr [1:130] "0" "1" "0" "0" ...
##  $ X19...Facial..Expression                                   : chr [1:130] "1" "1" "2" "2" ...
##  $ X20...Tremor..at..Rest.....head                            : chr [1:130] "0" "0" "0" "0" ...
##  $ X20...Tremor..at..Rest.....RUE                             : chr [1:130] "0" "0" "0" "1" ...
##  $ X20...Tremor..at..Rest.....LUE                             : chr [1:130] "2" "0" "0" "0" ...
##  $ X20...Tremor..at..Rest.....RLE                             : chr [1:130] "0" "0" "0" "1" ...
##  $ X20...Tremor..at..Rest.....LLE                             : chr [1:130] "2" "0" "0" "0" ...
##  $ X21...Action..or..Postural..Tremor.....RUE                 : chr [1:130] "0" "1" "0" "1" ...
##  $ X21...Action..or..Postural..Tremor.....LUE                 : chr [1:130] "0" "1" "0" "1" ...
##  $ X22...Rigidity.....neck                                    : chr [1:130] "0" "1" "2" "1" ...
##  $ X22...Rigidity.....RUE                                     : chr [1:130] "0" "0" "0" "1" ...
##  $ X22...Rigidity.....LUE                                     : chr [1:130] "1" "1" "0" "1" ...
##  $ X22...Rigidity.....RLE                                     : chr [1:130] "0" "1" "2" "1" ...
##  $ X22...Rigidity.....LLE                                     : chr [1:130] "0" "2" "2" "1" ...
##  $ X23.Finger..Taps.....RUE                                   : chr [1:130] "0" "1" "1" "1" ...
##  $ X23.Finger..Taps.....LUE                                   : chr [1:130] "1" "2" "1" "1" ...
##  $ X24...Hand..Movements.....RUE                              : chr [1:130] "0" "0" "0" "2" ...
##  $ X24...Hand..Movements.....LUE                              : chr [1:130] "0" "1" "1" "2" ...
##  $ X25...Rapid..Alternating..Movements.....RUE                : chr [1:130] "0" "1" "0" "1" ...
##  $ X25...Rapid..Alternating..Movements.....LUE                : chr [1:130] "1" "2" "0" "2" ...
##  $ X26...Leg..Agility.....RLE                                 : chr [1:130] "0" "1" "1" "1" ...
##  $ X26...Leg..Agility.....LLE                                 : chr [1:130] "0" "2" "2" "0" ...
##  $ X27...Arising..from..Chair                                 : chr [1:130] "0" "0" "0" "0" ...
##  $ X28...Posture                                              : chr [1:130] "0" "1" "3" "1" ...
##  $ X29...Gait                                                 : chr [1:130] "0" "0" "0" "1" ...
##  $ X30...Postural..Stability                                  : chr [1:130] "0" "1" "0" "0" ...
##  $ X31...Body..Bradykinesia..and..Hypokinesia                 : chr [1:130] "0" "1" "2" "1" ...
##  $ Entropy..of..speech..timing.....                           : num [1:130] 1.56 1.56 1.55 1.52 1.5
##  $ Rate..of..speech..timing.....min.                          : int [1:130] 354 340 211 140 269 31
##  $ Acceleration..of..speech..timing.....min2.                 : num [1:130] 6.05 27.52 11.97 -2.49
##  $ Duration..of..pause..intervals...ms.                       : int [1:130] 146 173 377 360 211 18
##  $ Duration..of..voiced..intervals...ms.                      : int [1:130] 264 253 322 663 328 28
##  $ Gaping..in.between..voiced..intervals.....min.             : num [1:130] 58.6 48.3 47.5 13.7 42
```

```
##  $ Duration..of..unvoiced..stops...ms.                          : num [1:130] 31.4 22.4 38.1 44.9 47
##  $ Decay..of..unvoiced..fricatives...â...min.                   : num [1:130] -2.101 -1.745 2.657 -0
##  $ Relative..loudness..of..respiration...dB.                    : num [1:130] -22.5 -24.6 -16.9 -25.5
##  $ Pause..intervals..per..respiration.....                      : num [1:130] 4.5 7 3 1 5 2.75 5.25 6
##  $ Rate..of..speech..respiration.....min.                       : num [1:130] 21.1 15.3 20.8 18.7 16
##  $ Latency..of..respiratory..exchange...ms.                     : int [1:130] 167 163 372 119 78 124
##  $ Entropy..of..speech..timing......1                           : num [1:130] 1.56 1.57 1.55 1.54 1.5
##  $ Rate..of..speech..timing.....min..1                          : int [1:130] 333 285 247 112 230 181
##  $ Acceleration..of..speech..timing.....min2..1                 : num [1:130] -2.82 8.2 4.71 -9.09 1
##  $ Duration..of..pause..intervals...ms..1                       : int [1:130] 158 295 280 397 206 611
##  $ Duration..of..voiced..intervals...ms..1                      : int [1:130] 318 264 317 800 480 398
##  $ Gaping..in.between..voiced..Intervals.....min.               : num [1:130] 49 40.6 49 18.7 33.5 .
##  $ Duration..of..unvoiced..stops...ms..1                        : num [1:130] 22.4 26.9 22.4 49.4 26
##  $ Decay..of..unvoiced..fricatives...â...min..1                 : num [1:130] 0.588 -0.825 -0.955 0.7
##  $ Relative..loudness..of..respiration...dB..1                  : num [1:130] -19.8 -23.3 -13.3 -25.1
##  $ Pause..intervals..per..respiration......1                    : num [1:130] 6 4 4 2 5 2 5 13.5 4 3
##  $ Rate..of..speech..respiration.....min..1                     : num [1:130] 13.8 21.8 22.5 14.4 14
##  $ Latency..of..respiratory..exchange...ms..1                   : int [1:130] 127 313 201 151 151 593
```

Como podemos ver, en el dataset original las variables tienen nombres muy largos y engorrosos por lo que primero las renombraremos, para poder trabajar con ellas con mas facilidad, y a continuación detallaremos la información que aporta cada una:

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0       0       0       0       0       0
```

El dataset consta de las siguientes variables para cada una de las observaciones:

code: (carácter) Contiene un código de identificación para cada paciente.

age: (numérica) Edad de cada paciente en años.

gender: (categórica con 2 niveles) Género del paciente.

history: (categórica de 2 niveles) Variable que indica si el paciente tiene familiares con Parkinson.

onset: (numérica) Edad del paciente al inicio de la enfermedad, en años.

duration: (numérica) Duración de la enfermedad desde los primeros síntomas, en años.

antidepr: (categórica con 10 niveles): Terapia con Antidepresivos del paciente, en caso afirmativo se especifica cuál.

antipark: (categórica con 1 nivel): Medicación antiparkinsoniana del paciente.

antipsych: (categórica con 1 nivel): Medicación antipsicótica del paciente.

benzodiazepine: (Categórica con 4 niveles): Medicación con benzodiazepinas del paciente, en caso afirmativo se especifica cuál.

levodopa: (numérica) Cantidad en miligramos del consumo de Levodopa diario del paciente.

clonazepam: (numérica) Cantidad en miligramos del consumo de Clonazepam diario del paciente.

Mediciones acústicas:

EST: (numérica) Mide la heterogeneidad en el habla en términos de la ocurrencia de intérvalos de sonoridad, insonoridad, pausas y respiraciraciones. (este valor se haciendo uso de la entropia de Shannon)

RST: (numérica) Mide la velocidad de elocución respecto a la calidad de ?¿?¿?¿

AST: (numérica) Mide la aceleración o deceleración en la velocidad del habla.

DPI: (numérica) Mide la media de la duración de los intérvalos de pausas de cada paciente.

DVI: (numérica) Mide la duración media de los intérvalos de voz.

GVI: (numérica) Separación media entre intérvalos de voz.

DUS: (numérica)

DUF: (numérica)

RLR: (numérica)

PIR: (numérica)

RSR: (numérica)

LRE: (numérica)