

Project Report - MAS8405

Lloyd Bates

25/02/2022

Data Description

Original Data

```
# First few entries of data  
head(as.data.frame(Reisby))
```

```
##      id hd week   lnimi   lndmi female reactive_depression  
## 1 101 23    0 4.04305 4.20469      0              1  
## 2 101 12    1 3.93183 4.81218      0              1  
## 3 101  9    2 4.33073 4.96284      0              1  
## 4 101  8    3 4.36945 4.96284      0              1  
## 5 103 13    0 2.77259 5.23644      1              1  
## 6 103 22    1 3.46574 5.20949      1              1
```

The data set contains 250 observations of 7 variables: “id”, “hd”, “week”, “lnimi”, “lndmi”, “female”, “reactive_depression”. Here one of the main variables that we will be observing is “lnimi” which represents the log concentration the antidepressant drug Imipramine (IMI) in a patients blood. The question of this report is observe the effectiveness of this drug on a patients depression.

Data Preprocessing

01-A File

Before analysis can be completed it is important to format the data correctly, one detail about the data set is that each row is not a unique person in total there are:

```
# Number of different patients  
groups = unique(reisby$id)  
length(groups)
```

```
## [1] 66
```

```
# Re-encode id vector  
reisby$id = match((reisby$id), groups)
```

This number means that each person was not measured every week. As well the id column has been modified to go from 1:66. Further data transformations have taken place including: converting the week column to a 0/1 representing a placebo week or a week where the drug was administered, also, the Hamilton depression index has been encoded to show the 4 possible levels of depression. Finally, for both a normalized and raw format, the data as been split into test and train data. Note groups are not split as `train[200]` is 53 and `test[201] = 54`

```
head(reisby)
```

```
##   id hd week   lnimi   lndmi female reactive_depression
## 1  1  2    0 4.04305 4.20469      0                1
## 2  1  1    0 3.93183 4.81218      0                1
## 3  1  1    0 4.33073 4.96284      0                1
## 4  1  1    0 4.36945 4.96284      0                1
## 5  2  1    0 2.77259 5.23644      1                1
## 6  2  2    0 3.46574 5.20949      1                1
```

Next the structure of the data frame can be viewed to see if how the variables are stored:

```
# View how data is stored
str(Reisby)
```

```
##  num [1:250, 1:7] 101 101 101 101 103 103 103 103 104 104 ...
##   - attr(*, "dimnames")=List of 2
##    ..$ : NULL
##    ..$ : chr [1:7] "id" "hd" "week" "lnimi" ...
```

Exploritory Data Analysis

Firstly to view any relationships between the variables the `sactterplotmatrix` can be viewed:

```
# View scatterplot matrix
pairs(reisby)
```

