Supervised methods

TASK 1 - Supervised methods: File `diet.dta` is a Stata database including information about several diseases and confounding variables (columns 1-16), nutrients (columns 17-26) and food consumption (columns 27-48).

1. Load data into R and save the information in an object called `diet`.
2. Create a train database selecting 4000 samples randomly and a test database with the rest by executing:

```
library(readstata13)
diet <- read.dta13("data_exercises/diet.dta")

vars <- !names(diet)%in%c("id", "casoc", "casom",
                          "casop", "casoe")
diet <- diet[ , vars]
diet <- diet[complete.cases(diet),]
diet$tipocancer <- droplevels(diet$tipocancer)


set.seed(12345)
sel <- sample(1:nrow(diet), 4000)
train <- diet[sel, ]
test <- diet[-sel,]
```

3. Perform a variable selection using Random Forest and Linear Discriminat methods using train dataset to predict the different types of cancer (variable `tipocancer`).
4. Evaluate model performance using the test dataset.