Analyse code COVID Quentin Duchemin

1. Importation des données, suppressions des valeurs impossibles (Nan)

* Pourquoi vecteurs à 2 lignes et 3 colonnes
* Pourquoi seq(22, 39)
* Pourquoi prediction c(2,3,40,42), seq(7,20)
* Comment sont les données sont elles pré-traitées ?

1. Prediction de l’infection
2. Etude des corrélations

* Corelation entre données 🡪 detection maladies et tous les paramètres pris en compte chez chaque patient
* Graphe= en fonction de l’intensité des couleurs, on sait avec quoi le fait d’être positif au covid est corrélé. Sur le schéma, on voit : plaquettes, leucocytes, eosinophiles, monocytes.

1. Utilité des prédicteurs

* Fonction glm : **glm** is used to fit generalized linear models, specified by giving a symbolic description of the linear predictor and a description of the error distribution.
* **glm.fit 1er argument :**  the result of a call to a family function
* Que veut dire .sep.function, .sep.glm 🡪 quel est le type de dfPrediction : tableau avec des vecteurs ?
* Comment savoir si les predicteurs ont une correlation positive avec le fait d’avoir le COVID ? pente de la courbe ?
* Likelihood Ratio test= calcul de la p valeur avec hypothèse nulle et hypothèse alternative ?

1. Régression logistique

* Division des données en deux parties : set d’entraînement et set de validation
* Floor 🡪 partie entière
* As.data.frame : donne un dataframe (c’est quoi) en sortie. -train\_ind 🡪 indices négatifs dans le dataframe ? permet de transformer une liste en un dataframe
* Cbind : ajouter une nouvelle colonne au dataframe rbind pour une nouvelle ligne
* deviance residuals **represent the contributions of individual samples to the deviance D**
* Pourquoi Moyenne negative ? Quelle unité ? Taux des composants dans le sang ?
* Roc curve= receiver operating characteristic curve, or ROC curve, is **a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied**
* Comment les predictions sur les données d’entrainement nous permettent elles d’avoir un modèle sur les données de test ?
* Training set / test data 🡪 lien?
* Matrice de confusion : connaître faux positifs et faux négatifs

What is a dataframe in R? [Managing Data Frames · UC Business Analytics R Programming Guide (uc-r.github.io)](http://uc-r.github.io/dataframes)

the easiest way to think of a data frame is as an Excel worksheet that contains columns of different types of data but are all of equal length rows

dfPrediction= dataframe for prediction

1. Trouver des maladies dont les symptômes sont similaires à ceux du covid