1. **CLASSIFICAZIONE TRAMITE REGRESSIONE LOGISTICA**

Carichiamo la tabella con anche la classe assegnata:

Ec <- read.table ('clipboard', header=TRUE)

Svolgiamo una classificazione non perentoria tramite il metodo della regressione logistica.

Rlog =glm(classe ~ Education+Health+Accidents+Employment+Export+HealthyYears+GenderPaymentGap+ IndustrialProduction+ TaxesLabour+ TaxesTotal, family = binomial, data=Ec)

oppure

Rlog =glm(classe ~., family = binomial, data=Ec)

Purtroppo fornisce errore. La cosa non è infrequente quando ci sono troppe variabili: si pensi che si chiede di minimizzare una funzione molto non lineare di 10 variabili. Un trucco per saltarci fuori è di ridurre la variabili e per farlo in fretta osserviamo la matrice di correlazione cor(Ec). Vediamo che la classe è più correlata a Employment, Export, TaxesTotal (se ne potrebbero aggiungere altre ma rischiamo di ricreare il problema).

Vediamo ora:

Rlog2 =glm(classe ~ Employment+Export+ TaxesTotal, family = binomial, data=Ec)

Ora ha funzionato. Vediamo i punteggi e le probabilità:

round(predict(Rlog2) ,3)

fornisce i punteggi del modello di regressione lieare multipla sottostante. Vediamo che i valori dell’output sono più accentuati, rispetto al caso della regressione fatta sopra. Ecco le relative probabilità:

round(predict(Rlog2,type="response"),3)

Il metodo ora è severissimo! Non dà scampo.

1. **Esercizi dai compiti d’esame**