

Analisi dei parametri vocali di pazienti affetti da malattia di Parkinson

Davide Dell'Orto - Matr. 828873

1. Descrizione ed esplorazione del dataset

```
##  subject age sex test_time motor_UPDRS total_UPDRS Jitter(%) Jitter(Abs)
Jitter.RAP Jitter.PPQ5
## 1      1  72  0   5.6431      28.199      34.398   0.00662   3.380e-05
0.00401    0.00317
## 2      1  72  0  12.6660      28.447      34.894   0.00300   1.680e-05
0.00132    0.00150
## 3      1  72  0  19.6810      28.695      35.389   0.00481   2.462e-05
0.00205    0.00208
##  Jitter.DDP Shimmer Shimmer(dB) Shimmer.APQ3 Shimmer.APQ5 Shimmer.APQ11
Shimmer.DDA      NHR
## 1    0.01204 0.02565      0.230      0.01438      0.01309      0.01662
0.04314 0.014290
## 2    0.00395 0.02024      0.179      0.00994      0.01072      0.01689
0.02982 0.011112
## 3    0.00616 0.01675      0.181      0.00734      0.00844      0.01458
0.02202 0.020220
##      HNR      RPDE      DFA      PPE
## 1 21.640 0.41888 0.54842 0.16006
## 2 27.183 0.43493 0.56477 0.10810
## 3 23.047 0.46222 0.54405 0.21014
```

Il dataset è composto da 5875 osservazioni e 22 variabili:

- *subject* - Codice dell'individuo (factor)
- *age* - Età dell'individuo (factor)
- *sex* - Sesso dell'individuo (factor)
- *test_time* - Giorni passati dal primo test (num)
- *motor_UPDRS* - Punteggio indicante l'intensità dei disturbi motori (num, da 0 a 108. I disturbi motori, includendo rigidità, tremore ed elasticità dei muscoli facciali, influenzano inevitabilmente il parlato)
- *total_UPDRS* - Punteggio indicante l'intensità della malattia (num, da 0 a 176 dove 176 rappresenta disabilità totale)
- Da *Jitter(%)* a *PPE* - Parametri vocali (num)

Non sono presenti *missing values*, infatti:

```
df[rowSums(is.na(df)) > 0, ]
```

```
## [1] subject      age      sex      test_time      motor_UPDRS
total_UPDRS
## [7] Jitter(%)      Jitter(Abs)    Jitter.RAP      Jitter.PPQ5      Jitter.DDP
Shimmer
## [13] Shimmer(dB)     Shimmer.APQ3    Shimmer.APQ5     Shimmer.APQ11    Shimmer.DDA      NHR
## [19] HNR              RPDE            DFA              PPE
## <0 righe> (o 0-length row.names)
```

Restituisce un data frame con 0 righe. A questo punto prendo in considerazione solo le variabili numeriche in modo da visualizzarne i principali indici descrittivi e calcolarne la matrice di correlazione:

```
numerics <- unlist(lapply(df, is.numeric))
dfnum <- df[, numerics]

mcor <- round(cor(dfnum), 2)
mcor[upper.tri(mcor)] <- ""
mcor <- as.data.frame(mcor)
mcor

##           age test_time motor_UPDRS total_UPDRS Jitter(%) Jitter(Abs)
Jitter.RAP Jitter.PPQ5
## age              1
## test_time      0.02          1
## motor_UPDRS    0.27      0.07          1
## total_UPDRS    0.31      0.08      0.95          1
## Jitter(%)      0.02     -0.02      0.08      0.07          1
## Jitter(Abs)    0.04     -0.01      0.05      0.07      0.87          1
## Jitter.RAP     0.01     -0.03      0.07      0.06      0.98      0.84
1
## Jitter.PPQ5    0.01     -0.02      0.08      0.06      0.97      0.79
0.95          1
## Jitter.DDP     0.01     -0.03      0.07      0.06      0.98      0.84
1      0.95
## Shimmer        0.1      -0.03      0.1      0.09      0.71      0.65
0.68      0.73
## Shimmer(dB)    0.11     -0.03      0.11      0.1      0.72      0.66
0.69      0.73
## Shimmer.APQ3   0.1      -0.03      0.08      0.08      0.66      0.62
0.65      0.68
## Shimmer.APQ5   0.09     -0.04      0.09      0.08      0.69      0.62
0.66      0.73
## Shimmer.APQ11  0.14     -0.04      0.14      0.12      0.65      0.59
0.6      0.67
## Shimmer.DDA    0.1      -0.03      0.08      0.08      0.66      0.62
0.65      0.68
## NHR            0.01     -0.03      0.07      0.06      0.83      0.7
0.79      0.86
## HNR            -0.1      0.04     -0.16     -0.16     -0.68     -0.71
-0.64     -0.66
## RPDE           0.09     -0.04      0.13      0.16      0.43      0.55
```

```

0.38      0.38
## DFA      -0.09      0.02      -0.12      -0.11      0.23      0.35
0.21      0.18
## PPE      0.12      0      0.16      0.16      0.72      0.79
0.67      0.66
##          Jitter.DDP Shimmer Shimmer(dB) Shimmer.APQ3 Shimmer.APQ5
Shimmer.APQ11 Shimmer.DDA
## age
## test_time
## motor_UPDRS
## total_UPDRS
## Jitter(%)
## Jitter(Abs)
## Jitter.RAP
## Jitter.PPQ5
## Jitter.DDP          1
## Shimmer          0.68          1
## Shimmer(dB)      0.69      0.99          1
## Shimmer.APQ3      0.65      0.98      0.97          1
## Shimmer.APQ5      0.66      0.98      0.98      0.96          1
## Shimmer.APQ11      0.6      0.94      0.94      0.89      0.94
1
## Shimmer.DDA      0.65      0.98      0.97          1      0.96
0.89          1
## NHR          0.79      0.8      0.8      0.73      0.8
0.71      0.73
## HNR          -0.64      -0.8      -0.8      -0.78      -0.79
-0.78      -0.78
## RPDE          0.38      0.47      0.47      0.44      0.45
0.48      0.44
## DFA          0.21      0.13      0.13      0.13      0.13
0.18      0.13
## PPE          0.67      0.62      0.64      0.58      0.59
0.62      0.58
##          NHR      HNR RPDE  DFA PPE
## age
## test_time
## motor_UPDRS
## total_UPDRS
## Jitter(%)
## Jitter(Abs)
## Jitter.RAP
## Jitter.PPQ5
## Jitter.DDP
## Shimmer
## Shimmer(dB)
## Shimmer.APQ3
## Shimmer.APQ5
## Shimmer.APQ11
## Shimmer.DDA
## NHR          1

```

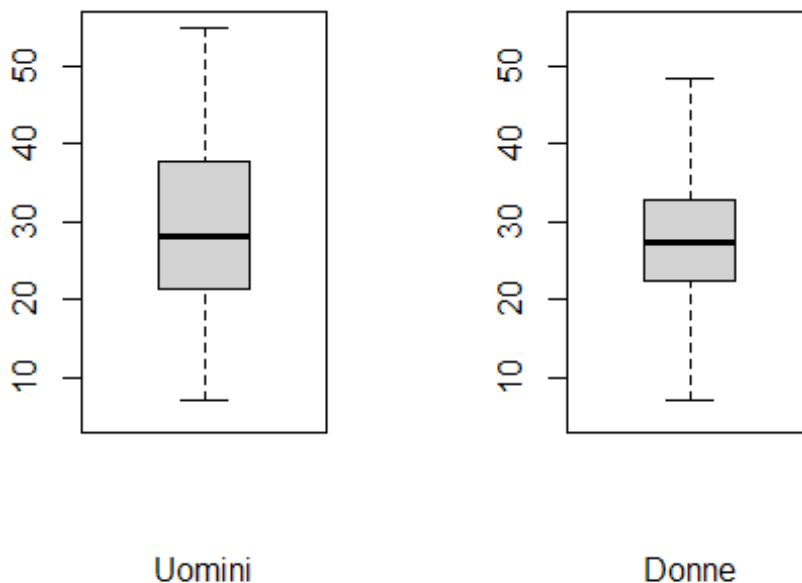
```
## HNR      -0.68      1
## RPDE      0.42 -0.66      1
## DFA      -0.02 -0.29 0.19      1
## PPE      0.56 -0.76 0.57 0.39      1
```

2. Inferenza su medie

Voglio verificare che esista una differenza tra i punteggi medi di uomo e donna. Precisamente, che la media di total_UPDRS, ovvero l'intensità media della malattia, sia significativamente diversa tra i due sessi.

```
uomini <- df[df$sex == "0", ]
donne <- df[df$sex == "1", ]

par(mfrow = c(1, 2))
boxplot(uomini$total_UPDRS, xlab = "Uomini", ylim = c(5, 55))
boxplot(donne$total_UPDRS, xlab = "Donne", ylim = c(5, 55))
```



```
mean(uomini$total_UPDRS)
## [1] 29.72406
mean(donne$total_UPDRS)
## [1] 27.50523
```

I dati sul campione suggeriscono che vi è una differenza di distribuzione tra gli uomini e le donne così come valori medi di total_UPDRS leggermente diversi. Verifico allora, tramite il test t, che

questa differenza in media non sia dovuta al caso ma sia da attribuirsi all'intera popolazione. In particolare, mi accerto prima della significatività della differenza nelle distribuzioni tramite il test chi-quadrato, ovvero mi accerto che la variabile `total_UPDRS` dipenda dal sesso:

```
UPDRS <- data.frame(c(uomini$total_UPDRS, donne$total_UPDRS), rep(c("M", "W"),
c(4008, 1867)))
colnames(UPDRS) <- c("total_UPDRS", "Sesso")

classes <- c(seq(5, 55, by = 5))
tab <- table(cut(df$total_UPDRS, classes), UPDRS$Sesso)

##
## Pearson's Chi-squared test
##
## data:  tab
## X-squared = 1260.2, df = 9, p-value < 2.2e-16
```

Dato un $p\text{-value} < 0.05$, rifiuto quindi l'ipotesi che le variabili `total_UPDRS` e `sex` siano indipendenti. Procedo allora a saggiare la significatività di questa differenza. Nello specifico, dato che le osservazioni a disposizione mi suggeriscono che negli uomini la media è superiore, voglio porre come ipotesi nulla $H_0: \text{mean_UPDRS}(\text{uomini}) \leq \text{mean_UPDRS}(\text{donne})$.

```
t.test(uomini$total_UPDRS, donne$total_UPDRS, var.equal = FALSE, alternative =
"greater")

##
## Welch Two Sample t-test
##
## data:  uomini$total_UPDRS and donne$total_UPDRS
## t = 7.74, df = 4032, p-value = 6.241e-15
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  1.747185      Inf
## sample estimates:
## mean of x mean of y
##  29.72406  27.50523
```

Con un $p\text{-value} < 0.05$ rifiuto l'ipotesi nulla e posso quindi dire che, in media, valori maggiori negli uomini non sono dovuti al caso, ma sono osservabili su un'ipotetica intera popolazione.

3. Regressione lineare multipla

Prima di procedere con l'individuazione dei regressori è necessario eliminare le variabili eccessivamente correlate tra di loro, al fine di evitare il problema della multicollinearità. Dato l'alto numero di variabili e l'importante dimensione della matrice di correlazione, creo una funzione `select` che mi permette di visualizzare rapidamente le correlazioni che superano una determinata soglia, in questo caso quelle maggiori (minori) di 0.8 (-0.8):

```
select <- function(x, value) {
  ind <- which(upper.tri(x), arr.ind = TRUE)
```

```

maxcor <- data.frame(X = dimnames(x)[[2]][ind[, 2]], Y =
dimnames(x)[[1]][ind[, 1]], corr = x[ind])
return(maxcor[abs(maxcor$corr) >= value, ])
}

```

```
select(cor(dfnum), 0.8)
```

```

##           X           Y      corr
## 6   Jitter(Abs)   Jitter(%) 0.8655772
## 9   Jitter.RAP   Jitter(%) 0.9841807
## 10  Jitter.RAP   Jitter(Abs) 0.8446263
## 13  Jitter.PPQ5   Jitter(%) 0.9682144
## 15  Jitter.PPQ5   Jitter.RAP 0.9471959
## 18  Jitter.DDP   Jitter(%) 0.9841835
## 19  Jitter.DDP   Jitter(Abs) 0.8446304
## 20  Jitter.DDP   Jitter.RAP 0.9999996
## 21  Jitter.DDP   Jitter.PPQ5 0.9472026
## 36  Shimmer(dB)   Shimmer 0.9923341
## 44  Shimmer.APQ3   Shimmer 0.9798280
## 45  Shimmer.APQ3   Shimmer(dB) 0.9680148
## 53  Shimmer.APQ5   Shimmer 0.9849043
## 54  Shimmer.APQ5   Shimmer(dB) 0.9763726
## 55  Shimmer.APQ5   Shimmer.APQ3 0.9627230
## 63  Shimmer.APQ11   Shimmer 0.9354568
## 64  Shimmer.APQ11   Shimmer(dB) 0.9363381
## 65  Shimmer.APQ11   Shimmer.APQ3 0.8856954
## 66  Shimmer.APQ11   Shimmer.APQ5 0.9389349
## 74  Shimmer.DDA   Shimmer 0.9798273
## 75  Shimmer.DDA   Shimmer(dB) 0.9680143
## 76  Shimmer.DDA   Shimmer.APQ3 1.0000000
## 77  Shimmer.DDA   Shimmer.APQ5 0.9627231
## 78  Shimmer.DDA   Shimmer.APQ11 0.8856941
## 81           NHR   Jitter(%) 0.8252937
## 84           NHR   Jitter.PPQ5 0.8648643
## 99           HNR   Shimmer -0.8014160
## 100          HNR   Shimmer(dB) -0.8024965

```

Per ogni coppia di variabili viene eliminata quella correlata in misura minore con la variabile target, ritrovandosi al termine con 8 variabili, di cui 7 potenzialmente esplicative:

```

##   age total_UPDRS Jitter(%) Shimmer.APQ11   HNR   RPDE   DFA   PPE
## 1  72      34.398  0.00662      0.01662 21.640 0.41888 0.54842 0.16006
## 2  72      34.894  0.00300      0.01689 27.183 0.43493 0.56477 0.10810
## 3  72      35.389  0.00481      0.01458 23.047 0.46222 0.54405 0.21014

```

Si procede a questo punto alla costruzione del modello: la prima variabile introdotta è quella con la più alta correlazione con la variabile target total_UPDRS:

```

mcor <- round(cor(dfnum, dfnum$total_UPDRS), 4)
mcor[upper.tri(mcor)] <- ""
mcor <- as.data.frame(mcor[-2, ])

```

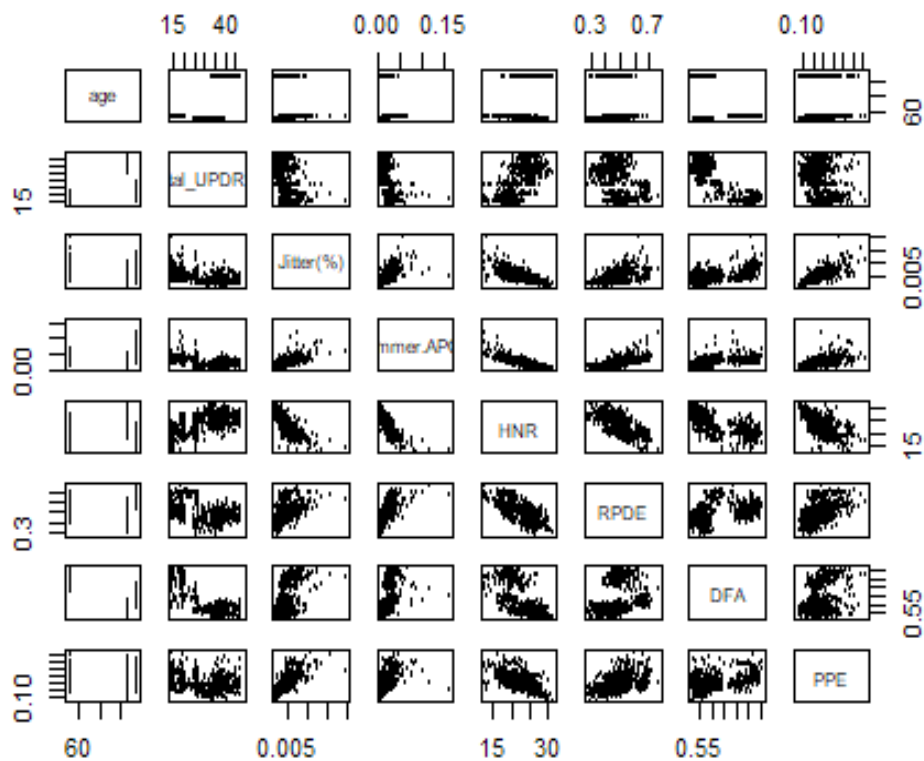
```

colnames(mcor) <- "total_UPDRS"
mcor

##          total_UPDRS
## age          0.3103
## Jitter(%)     0.0742
## Shimmer.APQ11  0.1208
## HNR           -0.1621
## RPDE           0.1569
## DFA           -0.1135
## PPE            0.1562

pairs(dfnum[sample(500), ], cex = 0.1)

```



Si inizializza quindi il modello con la variabile age:

```

model <- lm(df$total_UPDRS ~ df$age)
summary(model)

##
## Call:
## lm(formula = df$total_UPDRS ~ df$age)
##
## Residuals:
##      Min       1Q   Median       3Q      Max

```

```
## -22.092  -7.485  -1.191   7.730  23.641
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.62806    0.98409   4.703 2.62e-06 ***
## df$age       0.37637    0.01505  25.014 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.17 on 5873 degrees of freedom
## Multiple R-squared:  0.09628,    Adjusted R-squared:  0.09613
## F-statistic: 625.7 on 1 and 5873 DF,  p-value: < 2.2e-16
```

Entrambi i coefficienti risultano significativamente diversi da 0. L' R^2 è però ancora molto basso (0.096) ed è quindi necessario vedere se e come varia all'introduzione di nuovi regressori. Si sceglie allora di sviluppare nuovi modelli tramite *forward selection* dove ciascun modello include man mano una variabile esplicativa in più, fino da arrivare ad includerle tutte. Tramite il test ANOVA verrà scelto quel modello che avrà la più piccola somma del quadrato degli errori (RSS), per un determinato livello di significatività.

```
model1 <- update(model, . ~ . + df$`Jitter(%)`)
model2 <- update(model1, . ~ . + df$Shimmer.APQ11)
model3 <- update(model2, . ~ . + df$HNR)
model4 <- update(model3, . ~ . + df$RPDE)
model5 <- update(model4, . ~ . + df$DFA)
model6 <- update(model5, . ~ . + df$PPE)

anova(model, model1)

## Analysis of Variance Table
##
## Model 1: df$total_UPDRS ~ df$age
## Model 2: df$total_UPDRS ~ df$age + df$`Jitter(%)`
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1     5873 607797
## 2     5872 604768  1     3028.6 29.406 6.103e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(model, model2)

## Analysis of Variance Table
##
## Model 1: df$total_UPDRS ~ df$age
## Model 2: df$total_UPDRS ~ df$age + df$`Jitter(%)` + df$Shimmer.APQ11
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1     5873 607797
## 2     5871 603258  2     4539.1 22.088 2.775e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



```

anova(model, model3)

## Analysis of Variance Table
##
## Model 1: df$total_UPDRS ~ df$age
## Model 2: df$total_UPDRS ~ df$age + df$`Jitter(%)` + df$Shimmer.APQ11 +
##          df$HNR
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     5873 607797
## 2     5870 595329   3      12468 40.978 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(model, model4)

## Analysis of Variance Table
##
## Model 1: df$total_UPDRS ~ df$age
## Model 2: df$total_UPDRS ~ df$age + df$`Jitter(%)` + df$Shimmer.APQ11 +
##          df$HNR + df$RPDE
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     5873 607797
## 2     5869 593238   4      14559 36.009 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(model, model5)

## Analysis of Variance Table
##
## Model 1: df$total_UPDRS ~ df$age
## Model 2: df$total_UPDRS ~ df$age + df$`Jitter(%)` + df$Shimmer.APQ11 +
##          df$HNR + df$RPDE + df$DFA
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     5873 607797
## 2     5868 581455   5      26342 53.167 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(model, model6)

## Analysis of Variance Table
##
## Model 1: df$total_UPDRS ~ df$age
## Model 2: df$total_UPDRS ~ df$age + df$`Jitter(%)` + df$Shimmer.APQ11 +
##          df$HNR + df$RPDE + df$DFA + df$PPE
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     5873 607797
## 2     5867 576559   6      31238 52.98 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Il modello che implementa l’RSS minore è quello che include tutte le variabili.

```
##
## Call:
## lm(formula = df$total_UPDRS ~ df$age + df$`Jitter(%)` + df$Shimmer.APQ11 +
##     df$HNR + df$RPDE + df$DFA + df$PPE)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -25.377  -7.650  -1.637   7.413  28.318
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    25.96930     2.82358   9.197 < 2e-16 ***
## df$age          0.32379     0.01513  21.395 < 2e-16 ***
## df$`Jitter(%)` -148.58168    36.36661  -4.086 4.45e-05 ***
## df$Shimmer.APQ11 -29.09731    10.78713  -2.697 0.007008 **
## df$HNR          -0.30850     0.06343  -4.863 1.18e-06 ***
## df$RPDE          5.82108     1.73294   3.359 0.000787 ***
## df$DFA          -25.55140     2.02893 -12.594 < 2e-16 ***
## df$PPE           18.25037     2.58544   7.059 1.87e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.913 on 5867 degrees of freedom
## Multiple R-squared:  0.1427, Adjusted R-squared:  0.1417
## F-statistic: 139.5 on 7 and 5867 DF, p-value: < 2.2e-16
```

È evidente come il modello spieghi solamente poca parte della variabilità di total_UPDRS. L'R² aggiustato infatti, essendo pari a 0.142, non è migliorato di molto rispetto al modello base. Si è deciso allora di sviluppare un modello quadratico ovvero una regressione lineare multipla polinomiale. In particolare si è adottata una sorta di *backward selection* in quanto, partendo dal modello lineare *model6* sono stati fatti vari tentativi aggiungendo il quadrato di diverse variabili. Il modello scelto è il seguente:

```
model_b <- lm(df$total_UPDRS ~ df$age + df$`Jitter(%)` + df$Shimmer.APQ11 + df$HNR
+ df$DFA + df$RPDE +
  df$PPE + I(HNR^2), data = df)
##
## Call:
## lm(formula = df$total_UPDRS ~ df$age + df$`Jitter(%)` + df$Shimmer.APQ11 +
##     df$HNR + df$DFA + df$RPDE + df$PPE + I(HNR^2), data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -27.075  -7.331  -1.590   7.225  26.968
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.725894     3.407187   1.094  0.27420
## df$age          0.284043     0.015371  18.479 < 2e-16 ***
## df$`Jitter(%)` 106.653031    42.372688   2.517  0.01186 *
```

```
## df$Shimmer.APQ11 31.875732 11.936136 2.671 0.00759 **
## df$HNR           2.328663 0.239708 9.715 < 2e-16 ***
## df$DFA           -32.897634 2.107929 -15.607 < 2e-16 ***
## df$RPDE          4.267743 1.719617 2.482 0.01310 *
## df$PPE           12.747044 2.602662 4.898 9.96e-07 ***
## I(HNR^2)         -0.058866 0.005164 -11.399 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.806 on 5866 degrees of freedom
## Multiple R-squared:  0.1613, Adjusted R-squared:  0.1602
## F-statistic: 141 on 8 and 5866 DF, p-value: < 2.2e-16
```

Da cui risulta un'intercetta non significativamente diversa da zero. Si accetta quindi l'ipotesi alternativa secondo cui l'intercetta è pari a zero e la si elimina.

```
model_b <- lm(df$total_UPDRS ~ df$age + df$`Jitter(%)` + df$Shimmer.APQ11 + df$HNR
+ df$DFA + df$RPDE +
  df$PPE + I(HNR^2) - 1, data = df)

##
## Call:
## lm(formula = df$total_UPDRS ~ df$age + df$`Jitter(%)` + df$Shimmer.APQ11 +
##     df$HNR + df$DFA + df$RPDE + df$PPE + I(HNR^2) - 1, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -27.167  -7.328  -1.577   7.230  27.093
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## df$age           0.286994   0.015132  18.965 < 2e-16 ***
## df$`Jitter(%)`  125.179828  38.838925   3.223 0.001275 **
## df$Shimmer.APQ11 38.731032  10.157538   3.813 0.000139 ***
## df$HNR            2.517237   0.166505  15.118 < 2e-16 ***
## df$DFA          -32.351872   2.048030 -15.797 < 2e-16 ***
## df$RPDE           4.994946   1.585857   3.150 0.001643 **
## df$PPE           12.601901   2.599319   4.848 1.28e-06 ***
## I(HNR^2)         -0.062100   0.004233 -14.669 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.806 on 5867 degrees of freedom
## Multiple R-squared:  0.8996, Adjusted R-squared:  0.8995
## F-statistic: 6572 on 8 and 5867 DF, p-value: < 2.2e-16
```

L'R² aggiustato è aumentato esponenzialmente. È possibile cercare di semplificare il modello rimuovendo Jitter(%) ed eventualmente RPDE, vedendo se e quanto si modifica l'affidabilità:

```
##
## Call:
## lm(formula = df$total_UPDRS ~ df$age + df$Shimmer.APQ11 + df$HNR +
```

```
##      df$DFA + df$RPDE + df$PPE + I(HNR^2) - 1, data = df)
##
## Residuals:
##      Min        1Q    Median        3Q        Max
## -27.309  -7.391  -1.530    7.258   26.963
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## df$age          0.291066   0.015092   19.286 < 2e-16 ***
## df$Shimmer.APQ11 45.484972   9.946982    4.573 4.91e-06 ***
## df$HNR           2.308032   0.153455   15.040 < 2e-16 ***
## df$DFA          -31.004599   2.006520  -15.452 < 2e-16 ***
## df$RPDE           5.739757   1.570184    3.655 0.000259 ***
## df$PPE           16.747984   2.260495    7.409 1.45e-13 ***
## I(HNR^2)         -0.056663   0.003886  -14.582 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.814 on 5868 degrees of freedom
## Multiple R-squared:  0.8994, Adjusted R-squared:  0.8993
## F-statistic: 7497 on 7 and 5868 DF, p-value: < 2.2e-16
```

Si noti come, una volta rimossa la variabile Jitter(%), RPDE diventi significativamente diverso da zero e sia quindi opportuno mantenerlo all'interno del modello. Inoltre, l' R^2 è praticamente rimasto invariato e si ritiene così che questo possa essere il modello definitivo.