

Regresion Logistica

Cargar los datos

Cargamos los datos que se usaron durante el análisis descriptivo de los datos y mostramos las variables a usar.

```
datos_it0 <- read.csv("../data/bd_trabajo.csv")
names(datos_it0)

## [1] "id"
## [2] "diagnostico"
## [3] "d1"
## [4] "d2"
## [5] "d3"
## [6] "d4"
## [7] "d5"
## [8] "healthCode1"
## [9] "punto_medicacion"
## [10] "F0semitoneFrom27.5Hz_sma3nz_amean"
## [11] "F0semitoneFrom27.5Hz_sma3nz_stddevNorm"
## [12] "F0semitoneFrom27.5Hz_sma3nz_percentile20.0"
## [13] "F0semitoneFrom27.5Hz_sma3nz_percentile50.0"
## [14] "F0semitoneFrom27.5Hz_sma3nz_percentile80.0"
## [15] "F0semitoneFrom27.5Hz_sma3nz_pctlrange0.2"
## [16] "F0semitoneFrom27.5Hz_sma3nz_meanRisingSlope"
## [17] "F0semitoneFrom27.5Hz_sma3nz_stddevRisingSlope"
## [18] "F0semitoneFrom27.5Hz_sma3nz_meanFallingSlope"
## [19] "F0semitoneFrom27.5Hz_sma3nz_stddevFallingSlope"
## [20] "jitterLocal_sma3nz_amean"
## [21] "jitterLocal_sma3nz_stddevNorm"
## [22] "shimmerLocaldB_sma3nz_amean"
## [23] "shimmerLocaldB_sma3nz_stddevNorm"
## [24] "edad"
## [25] "cuidado"
## [26] "estimulacion_cerebral"
## [27] "anio_diagnostico"
## [28] "educacion"
## [29] "trabajo"
## [30] "genero"
## [31] "estado_civil"
## [32] "anio_medicacion"
## [33] "anio_enfermedad"
## [34] "facilidad_celular"
## [35] "fumo"
## [36] "cirugias"
```

```
## [37] "anios_fumo"
## [38] "healthCode"
```

No vamos a mostrar las medidas resumen o los gráficos de todas las variables ya que lo hicimos durante el análisis descriptivo. A lo largo del trabajo solo mostraremos los gráficos pertinentes.

Regresión Logística

En esta regresión vamos a utilizar como variable respuesta a **diagnostico**, usando las variables **d1**, **d2**, **d3**, **d4**, **d5** y **edad**. Para trabajar vamos a quedarnos únicamente con estas variables en el dataset.

```
datos_it1 <- datos_it0 %>%
  select(diagnostico, d1, d2, d3, d4, d5, edad)
names(datos_it1)
```

```
## [1] "diagnostico" "d1"          "d2"          "d3"          "d4"
## [6] "d5"          "edad"
```

Comenzamos realizando el modelo 0, donde se incluyen todas las variables posibles

```
modelo_it0 <- glm(diagnostico ~ ., data = datos_it1, family = "binomial")
summary(modelo_it0)
##
## Call:
## glm(formula = diagnostico ~ ., family = "binomial", data = datos_it1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.5698  -0.7436  -0.4146   0.8531   2.3973
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -8.513620   0.437964 -19.439  < 2e-16 ***
## d1           0.087467   0.012478   7.009 2.39e-12 ***
## d2          -0.003792   0.002220  -1.708  0.0876 .
## d3           0.220782   0.398317   0.554  0.5794
## d4           0.111645   2.591415   0.043  0.9656
## d5           0.757470   2.769715   0.273  0.7845
## edad        0.103631   0.004972  20.843  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 3023.1  on 2217  degrees of freedom
## Residual deviance: 2261.8  on 2211  degrees of freedom
## AIC: 2275.8
##
## Number of Fisher Scoring iterations: 4

r2 <- calcular_pseudo_R2(modelo_it0)
```

```

print("R2 del modelo: ")
## [1] "R2 del modelo: "
r2
## pseudo_R2    p_valor
## 0.2518205 0.0000000

```

Además, armaremos una tabla donde se especificara el AIC, R2 y alguna de observación de cada modelo armado, a manera comparación.

```

tabla_resumen_modelos <- data.frame(row.names = "modelo_it0",
                                     AIC = modelo_it0$aic,
                                     R2 = r2[1],
                                     Observacion = "Todas las variables")
rm(r2)

```