

INFORME PAC1

Taula de continguts

1. Abstract
 2. Objectius
 3. Materials i mètodes
 - 3.1. Origen i naturalesa de les dades
 - 3.2. Objecte *SummarizedExperiment*
 - 3.3. Preparació de dades
 - 3.4. Mètodes utilitzats
 4. Resultats
 - 4.1. Visualització de variables
 - 4.2. Anàlisi de relacions entre variables
 - 4.3. Correlacions
 - 4.4. Modelització
 5. Discussió, limitacions i conclusions
 - 5.1. Limitacions
 - 5.2. Recomanacions per a futurs estudis
-

Abstract

Aquest informe presenta l'anàlisi d'un *dataset* d'expressió òmica que inclou dades clíniques i metabolòmiques recollides de pacients. L'objectiu principal de l'estudi és explorar les interrelacions entre diverses variables clíniques i metabolòmiques i, a més, predir el valor de HBA1C a partir d'altres variables. S'ha utilitzat el paquet *SummarizedExperiment* per organitzar i manipular les dades. Les visualitzacions realitzades inclouen histogrames, boxplots, gràfics de dispersió i anàlisis de correlació. Finalment, s'ha ajustat un model de regressió lineal múltiple que ha proporcionat prediccions sobre HBA1C. Els resultats obtinguts proporcionen informació valuosa sobre les dinàmiques metabòliques i les interrelacions entre variables clau, contribuint així a la comprensió de factors de risc en condicions metabòliques.

Objectius

Els objectius principals d'aquesta PAC1 són:

- **Planificar i executar** una versió simplificada del procés d'anàlisi de dades òmiques, alhora practicar amb algunes de les eines i mètodes que s'han anat treballat al llarg de l'assignatura.
- **Explorar les dades clíniques i metabolòmiques:** Entendre la naturalesa de les dades i identificar patrons significatius.
- **Visualitzar la distribució de les variables:** Utilitzar histogrames i boxplots per identificar la distribució i la presència de valors atípics.

- **Analitzar alguna relació entre variables:** Crear gràfics de dispersió per observar les interrelacions entre les variables numèriques.
- **Calcular correlacions:** Examinar la matriu de correlació per identificar relacions significatives entre les variables.
- **Ajustar un model de regressió lineal múltiple:** Preveure el valor de HBA1C en funció d'altres variables independents i avaluar la seva significança.
- **Realitzar prediccions:** Aplicar el model ajustat per fer prediccions sobre els valors de HBA1C.

Materials i mètodes

1. Carregament de dades

El *dataset* seleccionat s'ha extret del repositori de github:

<https://github.com/nutrimetabolomics/metaboData/>

S'han utilitzat els paquets *readr* i *SummarizedExperiment* per carregar i manipular les dades. Els fitxers *DataInfo_S013.csv* i *DataValues_S013.csv* han estat carregats i explorats per extreure les metadades necessàries.

```
##{r}
# Carregar les llibreries
if (!requireNamespace("BiocManager", quietly = TRUE)) {
  install.packages("BiocManager")
}
BiocManager::install("SummarizedExperiment")
install.packages("readr")
library(SummarizedExperiment)
library(readr)
library(ggplot2)

# Carregar les dades
data_info <- read_csv("C:/Users/039421852J/Desktop/Laia/UOC/DadesOmiques/PAC1/DataInfo_S013.csv")
data_values <- read_csv("DataValues_S013.csv")

# Mostrar les primeres files per verificar que s'han carregat correctament
head(data_info)
head(data_values)
```

Les dades d'expressió òmica inclouen dues parts: *data_info*, que conté metadades sobre les mostres i les variables, i *data_values*, que inclou les mesures dels metabòlits i les dades clíniques dels pacients. És crucial revisar les primeres files de cada dataset per assegurar-nos que les dades s'han carregat correctament i per obtenir una idea general de les estructures i els continguts.

Figura 1. Head de *data_values*

A tibble: 6 x 696

...1 <dbl>	SUBJECTS <dbl>	SURGERY <chr>	AGE <dbl>	GENDER <chr>	Group <dbl>	MEDDM_T0 <dbl>	MEDCOL_T0 <dbl>	MEDINF_T0 <dbl>	MEDHTA_T0 <dbl>
1	1	by pass	27	F	1	0	0	0	1
2	2	by pass	19	F	2	0	0	0	0
3	3	by pass	42	F	1	0	0	0	0
4	4	by pass	37	F	2	0	0	0	0
5	5	tubular	42	F	1	0	0	0	0
6	6	by pass	24	F	2	0	0	0	0

6 rows | 1-10 of 696 columns

Figura 2. Head de data_info

...1 <chr>	VarName <chr>	varTpe <chr>	Description <chr>
SUBJECTS	SUBJECTS	integer	dataDesc
SURGERY	SURGERY	character	dataDesc
AGE	AGE	integer	dataDesc
GENDER	GENDER	character	dataDesc
Group	Group	integer	dataDesc
MEDDM_T0	MEDDM_T0	integer	dataDesc

6 rows

Per crear un objecte *SummarizedExperiment*, es va processar la matriu d'expressió i es van extreure les metadades de les mostres i variables.

```
{r}
# Crear matriu d'expressió amb els valors de les dades metabolòmiques
expression_data <- as.matrix(data_values[, -c(1:2)]) # Assumint que les primeres dues columnes són identificadors

# Metadades de les mostres
sample_metadata <- data_values[, 1:2] # Primeres dues columnes com metadades de les mostres

# Metadades dels metabòlits
variable_metadata <- data_info

# Crear l'objecte SummarizedExperiment
se <- SummarizedExperiment(
  assays = list(counts = expression_data),
  colData = sample_metadata,
  rowData = variable_metadata
)

# Comprovar l'objecte SummarizedExperiment creat
print(se)
```

La creació de l'objecte *SummarizedExperiment* permet gestionar les dades de manera eficient, associant les mesures d'expressió amb les metadades rellevants. Això és essencial per realitzar anàlisis que necessiten integrar dades clíniques i metabolòmiques. Les metadades de les mostres (com el sexe, l'edat i el grup de tractament) i les variables (descripcions dels metabòlits) són importants per interpretar els resultats i identificar possibles patrons.

2. Preparació de dades

Es va assegurar que les dades estiguessin en el format correcte, incloent la conversió de variables categòriques en factors i la gestió de valors nuls.

```
{r}
# Dimensions del dataset
dim(data_values)

# Tipus de dades de cada columna
str(data_values)

# Resum dels dades
summary(data_values)

# Convertir variables categòriques en factors
data_values$GENDER <- as.factor(data_values$GENDER)
data_values$Group <- as.factor(data_values$Group)
data_values$SURGERY <- as.factor(data_values$SURGERY)

# Comprovar si hi ha valors nuls
sum(is.na(data_values))
```

La funció `str()` és útil per entendre els tipus de dades de cada columna. Identificar les variables categòriques és fonamental per garantir que les anàlisis estadístiques es realitzin correctament. La gestió de valors nuls és una part crítica del preprocés, ja que poden afectar significativament els resultats de les anàlisis. La verificació de valors nuls ajuda a determinar si és necessari aplicar tècniques d'imputació o eliminar observacions incompletes.

S'ha observat que la dimensió de la base de dades és de 39:696 i conté un total de 3390 valors faltants.

3. Mètodes d'anàlisi utilitzats

- **Visualització de Dades**

En aquesta secció, l'objectiu és explorar les variables i comprendre millor la seva distribució i relacions. S'han utilitzat diversos tipus de gràfics per analitzar les dades:

Histogrames: Aquests gràfics mostren la distribució de freqüències d'una variable contínua. Cada barra representa la freqüència d'un interval de valors, la qual cosa permet identificar la forma de distribució de la variable (si és simètrica, asimètrica, etc.), així com la presència de valors extrems o anomalies.

Boxplots (diagrames de caixa i bigotis): Aquestes visualitzacions ofereixen una representació de la distribució de les dades per categories, mostrant estadístics com la mediana, els quartils, i els valors extrems. És especialment útil per identificar valors atípics i comparar la distribució d'una variable entre diferents grups o categories.

Gràfics de dispersió: S'han creat gràfics de dispersió per examinar la relació entre parells de variables numèriques. En aquests gràfics, cada punt representa una observació, la qual cosa permet detectar patrons de correlació o tendències entre dues variables. Per exemple, aquests gràfics poden ser útils per veure com la variable dependent (per exemple, HBA1C) varia en funció d'altres variables numèriques.

Aquest conjunt de visualitzacions proporciona una visió completa de les dades inicials, permetent identificar patrons i comportaments generals de les variables, així com relacions potencials que poden ser rellevants per a la modelització posterior.

- **Correlacions**

En aquesta secció, s'ha calculat la matriu de correlació per explorar les relacions entre les variables numèriques:

Matriu de correlació: La matriu de correlació mostra els coeficients de correlació entre cada parell de variables numèriques. Aquests coeficients, que oscil·len entre -1 i 1, indiquen el grau i la direcció de la relació lineal entre les variables. Valors propers a 1 o -1 indiquen una forta correlació positiva o negativa, respectivament, mentre que valors propers a 0 suggereixen una correlació feble o inexistent.

Per facilitar la interpretació, s'ha utilitzat el paquet *corrplot*, que genera una visualització gràfica de la matriu de correlació. Aquesta representació pot incloure cercles, colors i dimensions que indiquen la magnitud i la direcció de cada correlació, facilitant així la detecció de correlacions fortes i patrons interessants entre les variables.

- **Modelització**

En aquesta part, s'ha ajustat un model de regressió lineal múltiple amb l'objectiu d'estimar l'impacte de diverses variables en el valor d'HbA1C, que és una mesura del control del sucre en sang a llarg termini:

Ajust del model de regressió lineal múltiple: Aquest model permet analitzar l'efecte simultani de múltiples variables explicatives (predictors) sobre la variable dependent, HbA1C. S'estimen coeficients per a cada variable predictora, que representen el canvi esperat en el valor d'HbA1C per unitat de canvi en la variable predictora, mantenint la resta de variables constants.

Avaluació de l'impacte de les variables: A través dels coeficients del model, es pot interpretar quines variables tenen un efecte significatiu en els nivells de HbA1C. Això pot ajudar a identificar els factors més rellevants per al control de la glucosa en sang i guiar accions específiques o mesures preventives.

Aquesta secció proporciona una visió detallada de com les diferents variables influeixen en el resultat d'interès (HbA1C), cosa que pot resultar molt útil tant per a finalitats analítiques com per a la presa de decisions.

Resultats

Visualització de variables

S'han creat histogrames i boxplots per a les variables numèriques, permetent observar la seva distribució i detectar valors atípics.

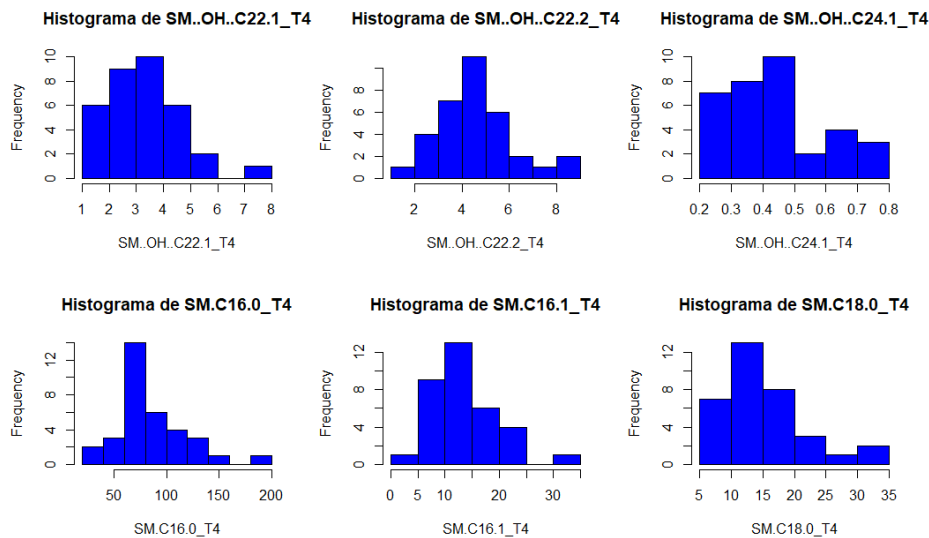
```
{r}
# Histogrames de les variables numèriques
numeric_columns <- sapply(data_values, is.numeric)
data_numeric <- data_values[, numeric_columns]

# Crear histogrames
par(mfrow=c(2, 3)) # Ajustem el layout
for (col in names(data_numeric)) {
  hist(data_numeric[[col]], main = paste("Histograma de", col), xlab = col, col = "blue", border = "black")
}

# Boxplots
par(mfrow=c(2, 3)) # Ajustem el layout
for (col in names(data_numeric)) {
  boxplot(data_numeric[[col]], main = paste("Boxplot de", col), ylab = col)
}
```

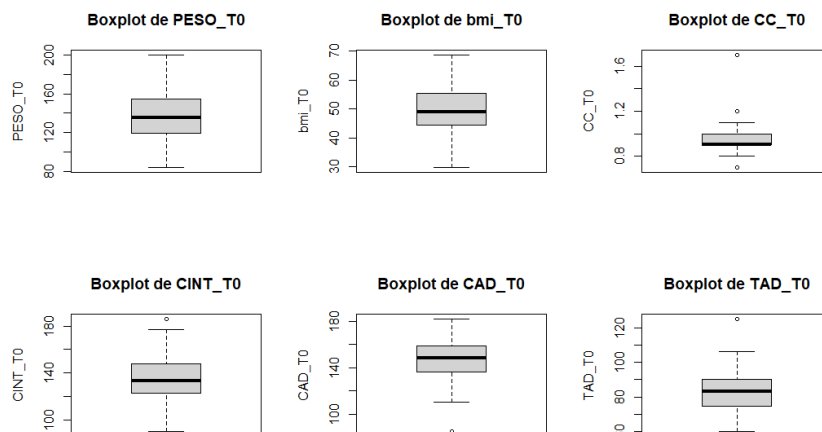
Histogrames: Permeten visualitzar la distribució de cada variable, ajudant a identificar si segueixen una distribució normal o si presenten skew (asimetria). La identificació de valors atípics a través dels histogrames és essencial per a la neteja de dades.

Gràfica 1. Exemple histogrames resultants d'algunes variables seleccionades



Boxplots: Són útils per comparar la variabilitat i detectar outliers (valors atípics) en les dades. Els valors atípics poden influir negativament en el model de regressió, per la qual cosa és important considerar la seva gestió.

Gràfica 2. Exemple boxplots resultants d'algunes variables seleccionades

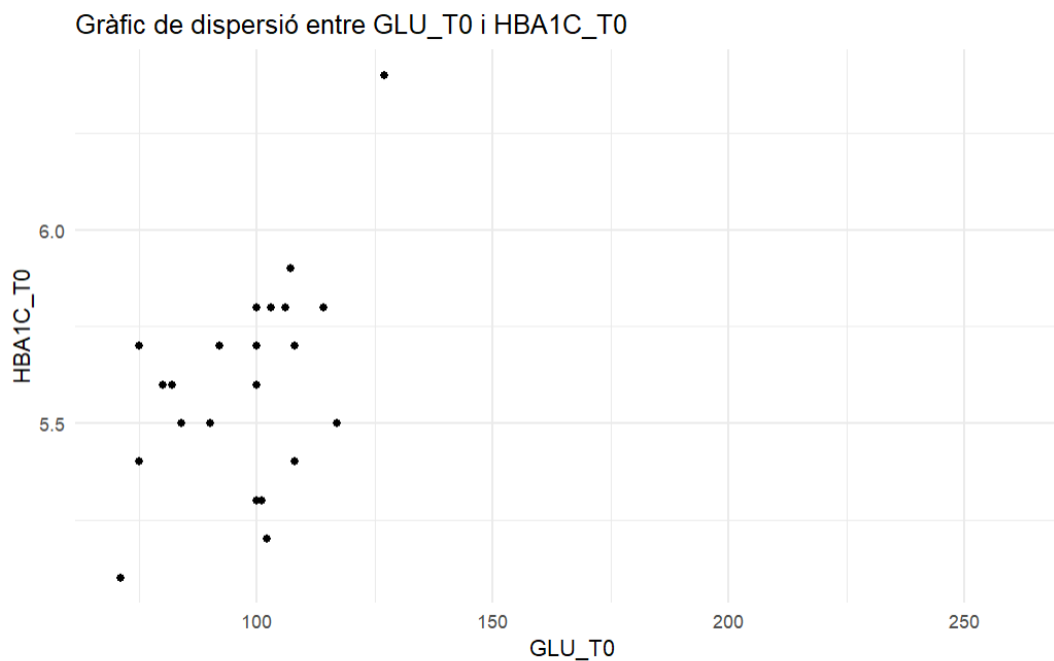


Anàlisi de relacions entre variables

S'ha creat un gràfic de dispersió per visualitzar la relació entre les variables GLU_T0 i HBA1C_T0.

```
{r}
# Gràfic de dispersió entre dues variables numèriques
ggplot(data_values, aes(x = GLU_T0, y = HBA1C_T0)) +
  geom_point() +
  labs(title = "Gràfic de dispersió entre GLU_T0 i HBA1C_T0", x = "GLU_T0", y = "HBA1C_T0") +
  theme_minimal()
```

Gràfica 3. Gràfic de dispersió



El gràfic de dispersió entre GLU_T0 i HBA1C_T0 permet visualitzar la relació entre els nivells de glucosa en dejú i els nivells de HBA1C. Una relació positiva entre aquestes dues variables podria suggerir que a mesura que augmenta el nivell de glucosa, també ho fa el nivell de HBA1C, un indicador clau del control de la glucosa a llarg termini.

Correlacions

La matriu de correlació s'ha calculat per a les variables numèriques i visualitzada mitjançant un gràfic de correlació.

```
```{r}
Matriu de correlació
cor_matrix <- cor(data_numeric, use = "complete.obs")

Visualitzar la matriu de correlació
library(corrplot)
corrplot(cor_matrix, method = "circle")
```
```

La matriu de correlació proporciona informació sobre la força i la direcció de les relacions lineals entre les variables. Correlacions fortes entre variables poden indicar que són interdependents, la qual cosa és rellevant a l'hora de construir models predictius. Per exemple, una correlació alta entre GLU_T0 i HBA1C_T0 suggereix que aquests dos metabòlits estan associats, cosa que és consistent amb el coneixement clínic.

Mostro, a continuació, una petita matriu de correlació entre algunes de les variables que tenim en el nostre *dataset*.

Figura 3. Resultats matriu correlació

| | MEDCOL_T0 | MEDINF_T0 | MEDHTA_T0 | GLU_T0 | INS_T0 |
|-----------|-------------|-------------|------------|-------------|-------------|
| MEDCOL_T0 | 1.00000000 | -0.06399222 | 0.2951052 | 0.05335307 | -0.10916003 |
| MEDINF_T0 | -0.06399222 | 1.00000000 | -0.2168454 | -0.11308150 | 0.01544078 |
| MEDHTA_T0 | 0.29510521 | -0.21684544 | 1.00000000 | 0.03123120 | 0.18153610 |
| GLU_T0 | 0.05335307 | -0.11308150 | 0.0312312 | 1.00000000 | 0.36756878 |
| INS_T0 | -0.10916003 | 0.01544078 | 0.1815361 | 0.36756878 | 1.00000000 |

En aquesta matriu de correlació, la majoria de relacions entre variables són febles o gairebé nul·les, indicant poca o cap relació lineal significativa. Només destaca la correlació moderada positiva entre GLU_T0 i INS_T0 (0.368), que suggereix una associació potencialment significativa entre els nivells de glucosa i insulina. La resta de correlacions són baixes i no revelen patrons clars entre les altres variables.

Modelització

S'ha ajustat un model de regressió lineal múltiple per preveure HBA1C_T0 en funció d'altres variables independents. Aquesta anàlisi ajuda a identificar quines variables tenen un impacte significatiu sobre HBA1C.

```
{r}
# Ajustar el model de regressió lineal múltiple
modelo <- lm(HBA1C_T0 ~ AGE + GENDER + PESO_T0 + bmi_T0 + GLU_T0 + INS_T0, data = data_values)

# Resum del model
summary(modelo)
```

```
Call:
lm(formula = HBA1C_T0 ~ AGE + GENDER + PESO_T0 + bmi_T0 + GLU_T0 +
    INS_T0, data = data_values)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.32991 -0.15074  0.04755  0.14100  0.43203
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.552537   0.484645   9.394 3.84e-08 ***
AGE          -0.012888   0.007372  -1.748 0.098447 .
GENDER        0.082123   0.182482   0.450 0.658372
PESO_T0       -0.007142   0.006453  -1.107 0.283866
bmi_T0         0.022389   0.017016   1.316 0.205722
GLU_T0         0.018291   0.004567   4.005 0.000917 ***
INS_T0        -0.023304   0.010781  -2.162 0.045205 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.2279 on 17 degrees of freedom
(15 observations deleted due to missingness)
Multiple R-squared:  0.5398,    Adjusted R-squared:  0.3774
F-statistic: 3.324 on 6 and 17 DF,  p-value: 0.02367
```

El resum del model proporciona informació crucial, incloent la R quadrada (que indica quina proporció de la variabilitat de HBA1C és explicada pel model) i els coeficients (que mostren l'impacte de cada variable independent). Variables amb un valor p significatiu indiquen que tenen un impacte estadísticament significatiu en HBA1C.

En aquesta anàlisi de regressió lineal múltiple, el model explora com diferents variables expliquen els nivells d'HBA1C_T0 (una mesura de glucosa en sang a llarg termini). A continuació es detalla la interpretació dels principals resultats:

1. Coeficients i significació estadística

- **Intercept (4.552537):** És el valor predit d'HBA1C_T0 quan totes les altres variables són zero. Aquest valor indica la contribució bàsica al valor d'HBA1C_T0.
- **AGE (-0.012888):** La variable AGE té un coeficient lleugerament negatiu, la qual cosa indica que, en general, a mesura que augmenta l'edat, l'HBA1C_T0 tendeix a disminuir, tot i que aquesta relació no és significativa (p-valor = 0.098).
- **GENDER (0.082123):** El coeficient per a GENDER és positiu, però amb un p-valor molt alt (0.658), la qual cosa indica que no és significatiu en aquest model.
- **PESO_T0 (-0.007142):** L'efecte del pes inicial és lleugerament negatiu, però tampoc és significatiu (p-valor = 0.283), indicant que el pes no té un efecte clar sobre HBA1C_T0.

- **bmi_T0 (0.022389)**: Tot i que el coeficient del BMI és positiu, la seva significació estadística és baixa (p-valor = 0.205), per la qual cosa tampoc contribueix significativament a predir HBA1C_T0.
- **GLU_T0 (0.018291)**: La glucosa inicial té un efecte positiu i significatiu (p-valor = 0.000917), indicant que l'augment de GLU_T0 s'associa amb un increment d'HBA1C_T0. Aquest és un dels factors amb més impacte en el model.
- **INS_T0 (-0.023304)**: La insulina inicial té un efecte negatiu i significatiu (p-valor = 0.045), cosa que suggereix que, a mesura que augmenta INS_T0, HBA1C_T0 tendeix a disminuir.

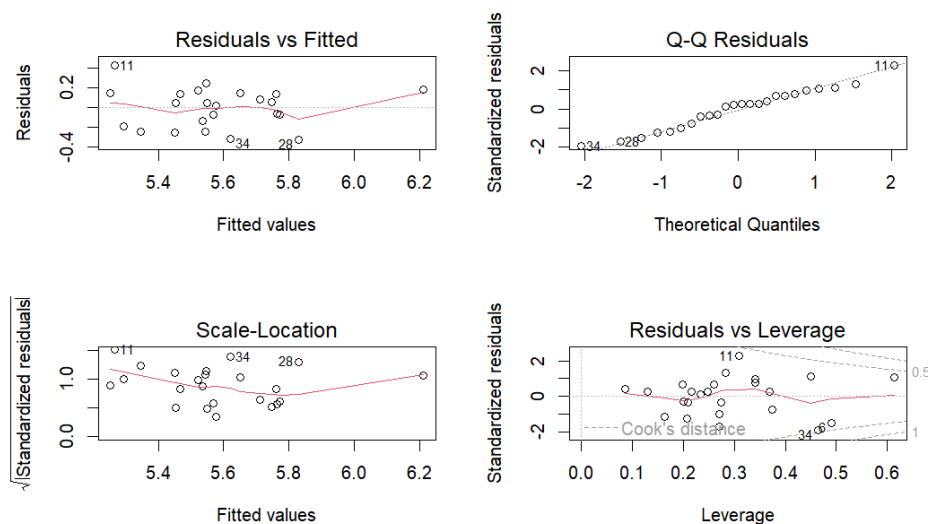
2. Qualitat del model

- **R-squared (0.5398) i Adjusted R-squared (0.3774)**: El model explica aproximadament el 54% de la variabilitat d'HBA1C_T0, i el R-squared ajustat (37.74%) indica que, si bé les variables explicatives ofereixen certa capacitat predictiva, hi ha factors addicionals fora del model que influeixen en HBA1C_T0.
- **F-statistic (3.324, p-value = 0.02367)**: Aquest valor indica que el model, en conjunt, és estadísticament significatiu (p-valor < 0.05), el que suggereix que almenys una de les variables predictors contribueix de manera significativa.

3. Residuals

Els valors dels residuals mostren com de lluny estan les observacions reals respecte a les prediccions. Aquests residuals es distribueixen principalment entre -0.33 i 0.43, amb una desviació estàndard de 0.2279, la qual cosa és moderadament ajustada per a un model predictiu.

Gràfica 4. Gràfiques per veure els residus del model



Prediccions

Finalment, s'han realitzat prediccions sobre HBA1C_T0 basades en el model ajustat.

```
## {r}
# Realitzar prediccions
prediccions <- predict(modelo, newdata = data_values)

# Afegir les prediccions al conjunt de dades
data_values$Prediccions_HBA1C_T0 <- prediccions

# Mostrar les primeres files amb les prediccions
head(data_values$Prediccions_HBA1C_T0)

|

```

| | 1 | 2 | 3 | 4 | 5 | 6 |
|--|----------|----------|----------|----------|----------|----------|
| | 5.823504 | 5.511969 | 5.254353 | 5.293637 | 5.550221 | 5.347256 |

Les prediccions realitzades permeten comparar els valors previstos amb els observats, facilitant l'avaluació del rendiment del model. Si el model s'ajusta bé a les dades, les prediccions haurien d'estar relativament a prop dels valors observats.

La utilització d'aquestes prediccions pot ser útil per identificar pacients en risc de desenvolupar complicacions relacionades amb la diabetis, permetent una intervenció precoç.

Discussió i conclusions

L'anàlisi realitzada ha proporcionat coneixements valuosos sobre les interrelacions entre les variables en el dataset d'expressió òmica.

Conclusions

En aquest estudi s'ha realitzat una anàlisi detallada de les dades amb l'objectiu de comprendre les relacions entre diverses variables i els seus efectes sobre els nivells d'HBA1C, un indicador clau del control de la glucosa en sang a llarg termini.

Els objectius de la PAC s'han assolit a través d'una planificació i execució simplificada del procés d'anàlisi de dades òmiques, utilitzant eines i mètodes treballats a l'assignatura, i s'ha completat amb èxit l'exploració de les dades clíniques i metabolòmiques per entendre la seva naturalesa i identificar patrons significatius. Les visualitzacions inicials, com ara histogrames, boxplots i gràfics de dispersió, han permès explorar la distribució de les variables i identificar possibles patrons o relacions preliminars, fet que ha facilitat la detecció de tendències i possibles valors atípics, especialment en les variables de glucosa i insulina, que posteriorment s'ha confirmat que tenen una rellevància significativa en l'anàlisi. En observar la matriu de correlació, s'ha detectat una associació moderada entre els nivells inicials de glucosa (GLU_T0) i insulina (INS_T0), suggerint que poden tenir una relació fisiològica important, mentre que les altres variables no han mostrat correlacions fortes entre elles, indicant que són principalment independents. Pel que fa al model de regressió lineal múltiple ajustat per predir HBA1C a partir de factors com l'edat, el gènere, el pes, l'IMC i els nivells inicials de glucosa i insulina, els resultats han determinat que GLU_T0 i INS_T0 són les úniques variables amb una contribució estadísticament significativa, amb una influència positiva de la primera i una influència negativa de la segona sobre HBA1C, la qual cosa subratlla la importància de la glucosa i la insulina com a predictors de l'HBA1C. Les altres variables, com edat, gènere, pes i IMC, no han mostrat un impacte significatiu en el model, la qual cosa suggereix que no tenen un paper rellevant en la predicció d'HBA1C en

aquesta mostra. Quant a la qualitat del model, malgrat ser estadísticament significatiu, només explica aproximadament el 54% de la variabilitat en els nivells d'HbA1C, fet que indica que altres factors no inclosos en el model podrien influir en aquest indicador, suggerint així la conveniència d'ampliar el model amb altres variables clíniques o comportamentals per tal de millorar la capacitat predictiva i obtenir una comprensió més completa dels factors que influeixen en l'HbA1C.

En conclusió, l'anàlisi indica que els nivells inicials de glucosa i insulina són factors significatius per explicar la variabilitat en HbA1C, i per tant, són elements clau a considerar en el monitoratge i la gestió del control glucèmic; no obstant això, el model presenta oportunitats de millora amb variables addicionals que podrien enriquir la comprensió de tots els factors que impacten en els nivells d'HbA1C, cosa que pot ser de gran utilitat per a professionals de la salut en el desenvolupament d'estratègies més efectives per a la prevenció i el control de la diabetis i altres afeccions metabòliques. A més, s'han obtingut prediccions sobre els valors de HbA1C en base al model ajustat, avaluant la seva capacitat per explicar part de la variabilitat d'aquest indicador i complint així amb els objectius plantejats de fer prediccions amb el model construït.

Limitacions de l'Estudi

Aquest estudi pot presentar diverses limitacions que comentaré per a una interpretació més acurada dels resultats.

En primer lloc, la capacitat predictiva del model és limitada, ja que només explica aproximadament el 54% de la variabilitat en els nivells d'HbA1C. Això indica que altres factors no inclosos, possiblement dades òmiques o variables clíniques addicionals, podrien influir significativament en l'HbA1C. A més, algunes variables incloses, com l'edat, el gènere, el pes i l'IMC, no han mostrat un impacte estadísticament significatiu, la qual cosa pot indicar una falta de representativitat en les dades o la necessitat de variables més rellevants per a explicar HbA1C. També, la mostra utilitzada en aquest estudi és reduïda, ja que es van eliminar observacions amb valors mancats, cosa que pot afectar la robustesa i la generalització dels resultats. A nivell metodològic, l'ús d'un model lineal pot limitar la detecció de relacions no lineals entre les variables; per tant, altres enfocaments, com els models no lineals o de machine learning, podrien oferir una millor capacitat predictiva i un ajust més detallat a les dades. Finalment, el model no té en compte variables potencialment rellevants com els hàbits dietètics, l'activitat física o altres factors ambientals, els quals podrien tenir una influència important sobre els nivells d'HbA1C i podrien millorar la comprensió global dels factors que afecten aquest indicador.

Aquestes limitacions suggereixen que una ampliació de l'estudi amb més dades, variables addicionals i metodologies més complexes podria proporcionar resultats més robustos i significatius per a l'anàlisi de HbA1C.

Repositori de GitHub

Per accedir al codi i a les dades utilitzades en aquest estudi, podeu consultar el repositori de GitHub a la següent direcció: <https://github.com/Laias6/Sola-Reguant-Laia-PEC1>. Aquí hi trobareu tots els scripts de preprocesament, anàlisi i visualització, així com la documentació necessària per reproduir els resultats de l'informe.