Actividad\_3. Análisis de un caso práctico en R

**TABLAS**

***Tabla 1. Varianza acumulada explicada por cada componente principal.***

|  |  |
| --- | --- |
|  | Varianza\_acumulada |
| Dim.1 | 52.5 |
| Dim.2 | 59 |
| Dim.3 | 64.3 |
| Dim.4 | 68.6 |
| Dim.5 | 72.3 |
| Dim.6 | 75.8 |
| Dim.7 | 78.8 |
| Dim.8 | 81.4 |
| Dim.9 | 83.8 |
| Dim.10 | 85.8 |
| Dim.11 | 87.6 |
| Dim.12 | 89.3 |
| Dim.13 | 90.6 |
| Dim.14 | 91.9 |
| Dim.15 | 92.9 |
| Dim.16 | 93.8 |
| Dim.17 | 94.6 |
| Dim.18 | 95.4 |
| Dim.19 | 96 |
| Dim.20 | 96.7 |
| Dim.21 | 97.1 |
| Dim.22 | 97.5 |
| Dim.23 | 97.8 |
| Dim.24 | 98.2 |
| Dim.25 | 98.4 |
| Dim.26 | 98.7 |
| Dim.27 | 98.9 |
| Dim.28 | 99 |
| Dim.29 | 99.2 |
| Dim.30 | 99.3 |
| Dim.31 | 99.4 |
| Dim.32 | 99.5 |
| Dim.33 | 99.6 |
| Dim.34 | 99.7 |
| Dim.35 | 99.7 |
| Dim.36 | 99.8 |
| Dim.37 | 99.8 |
| Dim.38 | 99.9 |
| Dim.39 | 99.9 |
| Dim.40 | 99.9 |
| Dim.41 | 100 |
| Dim.42 | 100 |
| Dim.43 | 100 |
| Dim.44 | 100 |
| Dim.45 | 100 |
| Dim.46 | 100 |

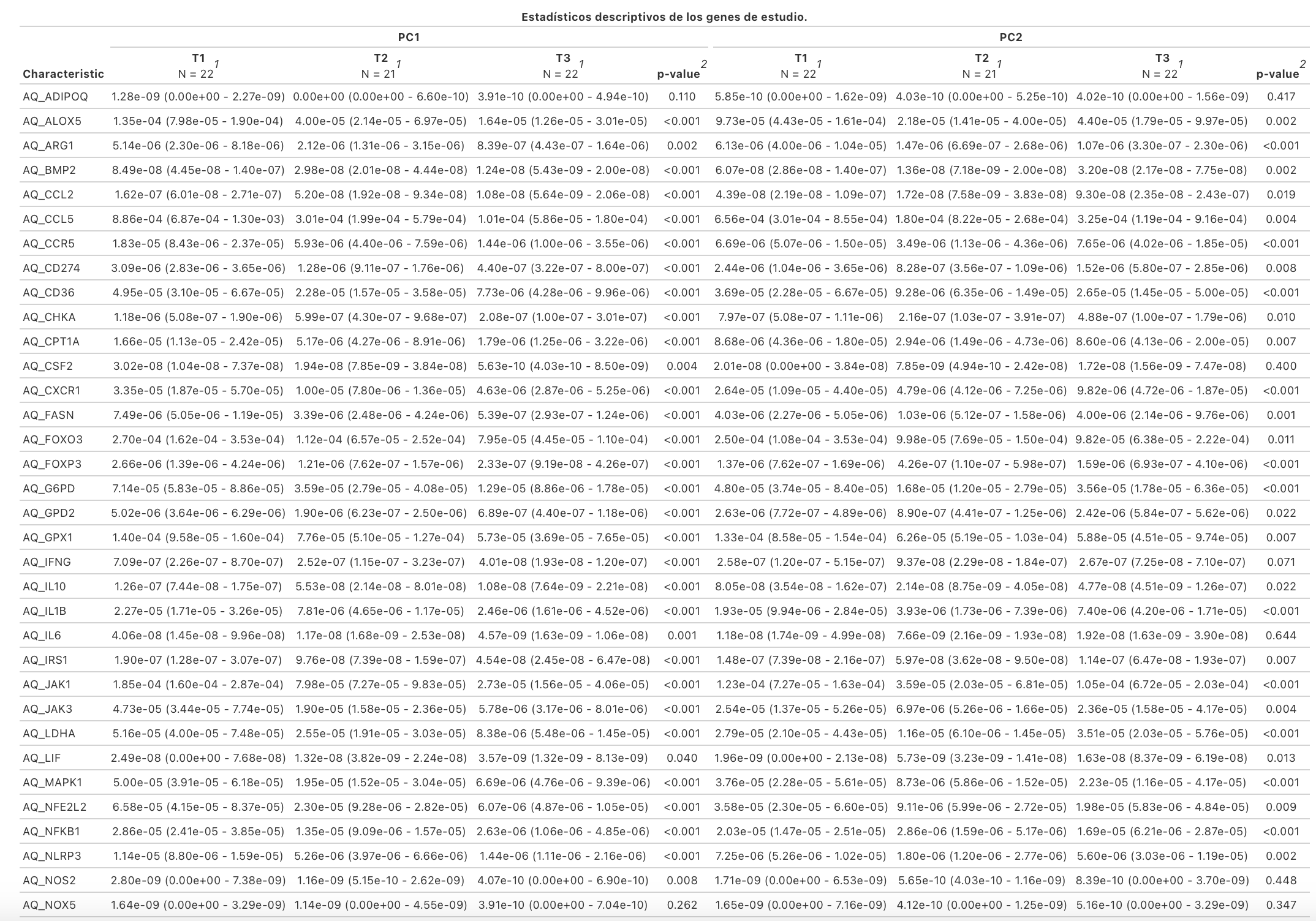
La Tabla 1 muestra el porcentaje acumulado de la varianza total explicada por cada uno de los 46 componentes estudiados en el Análisis de Componentes Principales (PCA) aplicado en el estudio de los genes.

***Tabla 2. Cargas de los genes para cada una de las primeras cinco dimensiones del PCA***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Dim.1 | Dim.2 | Dim.3 | Dim.4 | Dim.5 |
| AQ\_ADIPOQ | 0.04 | 0.06 | -0.2 | 0.62 | -0.12 |
| AQ\_ALOX5 | -0.16 | -0.09 | -0.02 | 0.01 | -0.16 |
| AQ\_ARG1 | -0.06 | -0.37 | 0.09 | 0 | 0.04 |
| AQ\_BMP2 | -0.12 | -0.1 | 0.02 | 0.05 | 0.3 |
| AQ\_CCL2 | -0.11 | 0.34 | -0.15 | -0.14 | -0.08 |
| AQ\_CCL5 | -0.18 | 0.03 | 0.08 | 0.06 | 0.05 |
| AQ\_CCR5 | -0.17 | 0.14 | 0.09 | 0.07 | -0.08 |
| AQ\_CD274 | -0.16 | -0.11 | -0.03 | 0.02 | 0.04 |
| AQ\_CD36 | -0.16 | -0.02 | 0.05 | 0.02 | -0.09 |
| AQ\_CHKA | -0.11 | 0.13 | 0.28 | 0.12 | 0.12 |
| AQ\_CPT1A | -0.17 | 0.16 | 0.11 | 0.01 | -0.13 |
| AQ\_CSF2 | -0.1 | 0.07 | -0.05 | -0.03 | 0.39 |
| AQ\_CXCR1 | -0.14 | -0.19 | -0.15 | -0.01 | 0.14 |
| AQ\_FASN | -0.18 | 0.17 | 0.03 | 0.04 | 0.09 |
| AQ\_FOXO3 | -0.12 | -0.08 | 0.28 | 0.04 | -0.2 |
| AQ\_FOXP3 | -0.14 | 0.26 | -0.14 | -0.15 | -0.07 |
| AQ\_G6PD | -0.17 | -0.12 | 0.18 | 0.07 | -0.12 |
| AQ\_GPD2 | -0.16 | 0.16 | 0.08 | 0.09 | -0.03 |
| AQ\_GPX1 | -0.13 | -0.1 | 0.15 | -0.09 | -0.02 |
| AQ\_IFNG | -0.16 | 0.03 | -0.01 | 0.1 | 0.07 |
| AQ\_IL10 | -0.1 | -0.27 | -0.23 | 0.02 | -0.05 |
| AQ\_IL1B | -0.15 | -0.17 | -0.21 | -0.04 | 0.12 |
| AQ\_IL6 | -0.12 | -0.04 | -0.05 | 0.07 | 0.43 |
| AQ\_IRS1 | -0.14 | 0.09 | 0.11 | -0.09 | -0.09 |
| AQ\_JAK1 | -0.19 | 0.08 | -0.01 | -0.01 | 0.1 |
| AQ\_JAK3 | -0.14 | -0.17 | -0.32 | -0.04 | 0.01 |
| AQ\_LDHA | -0.17 | 0.06 | 0.03 | 0.01 | -0.01 |
| AQ\_LIF | -0.12 | 0.31 | 0.07 | 0 | 0.09 |
| AQ\_MAPK1 | -0.18 | -0.09 | 0.1 | 0.05 | -0.03 |
| AQ\_NFE2L2 | -0.18 | -0.02 | 0.02 | 0 | 0.01 |
| AQ\_NFKB1 | -0.18 | 0 | 0.03 | 0.04 | 0.11 |
| AQ\_NLRP3 | -0.17 | 0.05 | -0.1 | -0.06 | -0.17 |
| AQ\_NOS2 | -0.1 | 0.07 | -0.27 | -0.19 | -0.33 |
| AQ\_NOX5 | 0.04 | 0.06 | -0.2 | 0.62 | -0.12 |
| AQ\_PDCD1 | -0.15 | 0.12 | 0.04 | 0.11 | -0.12 |
| AQ\_PPARG | -0.15 | -0.17 | 0.05 | 0.09 | -0.22 |
| AQ\_PTAFR | -0.18 | -0.15 | 0 | 0.01 | -0.01 |
| AQ\_PTGS2 | -0.15 | -0.06 | 0.16 | 0.08 | 0.05 |
| AQ\_SLC2A4 | -0.02 | -0.25 | 0.34 | -0.03 | -0.24 |
| AQ\_SOD1 | -0.16 | 0.1 | 0.08 | -0.02 | 0.01 |
| AQ\_SREBF1 | -0.18 | 0.11 | 0.01 | -0.03 | 0.06 |
| AQ\_STAT3 | -0.16 | 0.05 | -0.26 | -0.14 | -0.11 |
| AQ\_TGFB1 | -0.18 | -0.02 | -0.09 | -0.02 | -0.02 |
| AQ\_TLR3 | -0.15 | -0.05 | 0.02 | 0.16 | 0.13 |
| AQ\_TLR4 | -0.16 | -0.17 | -0.18 | -0.05 | -0.04 |
| AQ\_TNF | -0.16 | -0.06 | -0.15 | -0.06 | -0.06 |

En esta tabla se muestra el aporte y la dirección de la asociación de cada gen estudiado en los cinco primeros componentes en el Análisis de Componentes Principales (PCA). Un valor positivo indica una asociación directa, mientras que un valor negativo indica asociación inversa.

***Tabla 3. Estadísticas descriptivas de los genes de estudio.***



En la Tabla 3 se muestra la mediana y el rango intercuartílico de la expresión de cada gen para los pacientes pertenecientes al primer (T1), segundo (T2) y tercer (T3) tercil de las dos primeras dimensiones del PCA (PC1 y PC2). La clasificación en terciles se hizo mediante el valor del componente para cada paciente. La N indica el número de pacientes en cada tercil. Para comprobar diferencias de expresión significativas entre los terciles, se realizó un test Kruskal-Wallis para cada gen. Un p-valor<0.05 indicaría que uno o más de los grupos tiene una expresión significativamente distinta al resto.

***Tabla 4. Regresión logística binomial.***

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Categorías | OR (IC 95%) | P value |
| PC1\_tercil | T1 | 1 (Ref.) | NA |
|  | T2 | 0,4 (0,067 - 2,104) | 0,286 |
|  | T3 | 1,1 (0,138 – 8,423) | 0,930 |
| PC2\_tercil | T1 | 1 (Ref.) | NA |
|  | T2 | 2,988 (0,421 – 25,782) | 0,288 |
|  | T3 | 1,167 (0,237 – 5,999) | 0,848 |
| Vómitos | No | 1 (Ref.) | NA |
|  | Sí | 0,988 (0,055 – 27,724) | 0,993 |
| Tumor | CCR | 1 (Ref.) | NA |
|  | CM | 2,729 (0,685 – 12,602) | 0,169 |
|  | CP | 2,309 (0,346 – 18,441) | 0,400 |
| Neuropatía | No | 1 (Ref.) | NA |
|  | Sí | 4,391 (0,672 – 41,688) | 0,147 |
| IgG | - | 0,999 (0,997 – 1,001) | 0,289 |
| PCR | - | 1,042 (0,995 – 1,128) | 0,215 |
| Linfocitos | - | 1,518 (0,495 – 5,191) | 0,478 |
| IgE | - | 0,999 (0,990 – 1,009) | 0,820 |

La tabla recoge los valores de Odds Ratio (OR) con sus correspondientes intervalos de confianza al 95% (IC) de las diferentes variables incluidas en el modelo de regresión, así como el p-valor de los coeficientes. Para las variables categóricas PC1\_tercil, PC2\_tercil, Vómitos, Tumor y Neuropatía se tomó una de las categorías como referencia. Un valor de OR=1 indica que la variable no tiene riesgo o protección asociado a pertenecer al grupo de “metástasis: si”.

**INFORME**

Tras haber analizado el dataset, el cual contiene datos de 65 pacientes incluyendo información de la expresión de 46 genes y otras variables clínicas, hemos podido estudiar relaciones entre dichas variables y el riesgo asociado a tener metástasis.

En primer lugar, se realizó un análisis exhaustivo basado en los componentes principales (PCA) sobre la expresión de los 46 genes. En dicho análisis, se redujo la información con la que se iba a trabajar y se seleccionaron los dos componentes principales más importantes (PC1 y PC2), que explicaban más del 50% de la varianza de los datos. En las representaciones gráficas de los componentes principales, podemos destacar que la mayoría de los genes estaban bien representados por esas dos dimensiones.

A partir de los valores de cada paciente para cada el primer y segundo componente, se clasificó a los pacientes en tres terciles y se comparó la expresión mediana de cada uno de los 46 genes entre los distintos grupos generados. Con esto podemos corroborar que cada componente principal capta distintas diferencias (por ejemplo, entre los terciles del PC1 hay diferencias de expresión para IL6, pero no las hay entre los terciles de PC2; lo contrario también es cierto para SLC2A4).

Seguidamente, quisimos comprobar si pertenecer a alguno de estos terciles podría estar asociado a un riesgo mayor o menor de tener metástasis mediante una regresión logística binomial. No obstante, primero intentamos buscar otras variables de ajuste entre las variables clínicas o sociodemográficas del dataset original. Para ello intentamos buscar variables confusoras que pudieran estar asociadas a la variable de metástasis o a las variables de los terciles. Encontramos siete variables con cierta asociación a los grupos de terciles y decidimos añadirlas también al modelo, aunque la mayoría perdió la significancia en el valor p cuando lo corregimos por múltiples comparaciones.

Tanto el modelo inicial sin ajustar, como el modelo completo ajustado por variables clínicas, no mostraron ninguna asociación significativa. En ambos casos, los valores p fueron >0.05 y los IC de los Odd Ratio contenían el valor 1, lo cual indica que no hay una relación clara entre las variables independientes y la variable dependiente metástasis.

Podemos concluir pues, que pese a tener en cuenta la expresión génica de 46 genes (reduciendo su complejidad mediante PCA) y siete variables clínicas extra, no se han encontrado relaciones estadísticamente significativas con la presencia de metástasis. Esto podría deberse a que las variables consideradas no han sido las más relevantes para este resultado, a que no hemos considerado suficientes dimensiones en el PCA, o a que simplemente no hay una asociación real entre estas variables y tener o no metástasis.