

Capítulo 2

Caracterización estadística

Para realizar la caracterización estadística de los datos se han usado representaciones estadísticas clásicas que nos muestren de una manera visual las diferencias entre los grupos de estudio; así se ha utilizado diagramas de cajas y bigotes (Apartado 2.1) para resaltar la variabilidad estadística de las muestras. Para representar más claramente las diferencias se ha optado por el uso de histogramas (Apartado 2.2), que además nos aportan información sobre las posiciones BOCs más dominantes en cada uno de los grupos de estudio y la aportación menos significativa de las bandas en menores posiciones.

2.1. Diagrama de cajas y bigotes

El diagrama de cajas es una representación usada en estadística descriptiva y nos muestra de forma gráfica los valores estadísticos más importantes de un conjunto de datos agrupados en categorías. Así pues, la caja está acotada por los cuartiles Q1 y Q3 de las muestras, mientras que los bigotes representan los valores mínimos y máximos muestrales. En este tipo de diagramas también se muestran como puntos externos a la caja los datos considerados como outliers.

En la representación de cajas y bigotes de nuestros datos en la Figura 2.1, se puede apreciar la variabilidad estadística en las posiciones BOCs de ambos grupos de casos. Así, vemos que en ambos grupos, la distribución de la caja no es completamente simétrica y tienden a tener una mayor tamaño antes de sus respectivas medianas. No obstante, a pesar de esta similitud, los casos que convierten muestran una menor variabilidad estadística

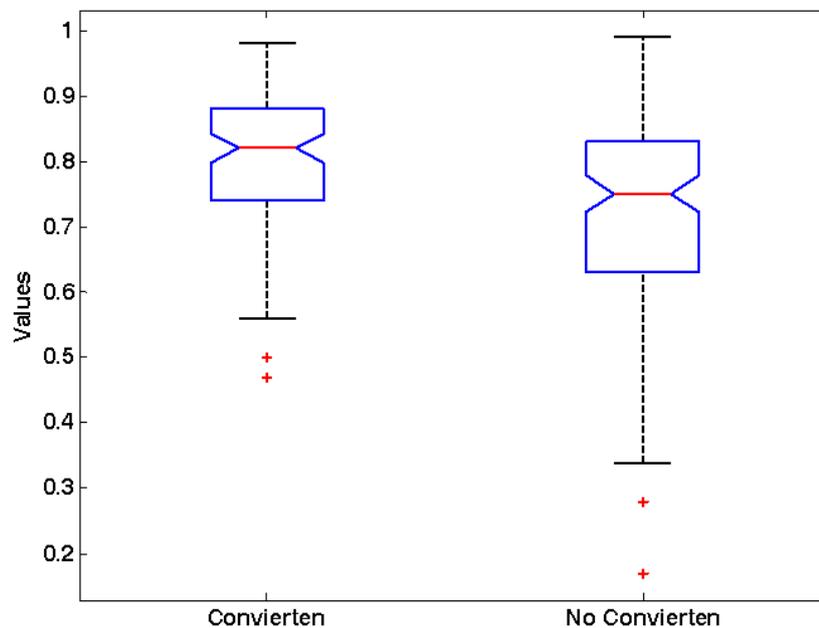


Figura 2.1: Representación de cajas y Bigotes de la agrupación de BOCs para casos que Convierten y que No Convierten.

y poseen la mediana de sus muestras en un punto superior al valor 0.8 de las posiciones, mientras que los casos que no convierten están mucho más dispersos y tienen su mediana antes de llegar a la citada posición.

Podemos apreciar también en nuestro diagrama que hay algunas posiciones consideradas como outliers. El criterio para considerar un punto como outlier viene definido por un factor de bigote $w = 1,5$, de forma que se encuentran aquellas posiciones, p , que cumplen:

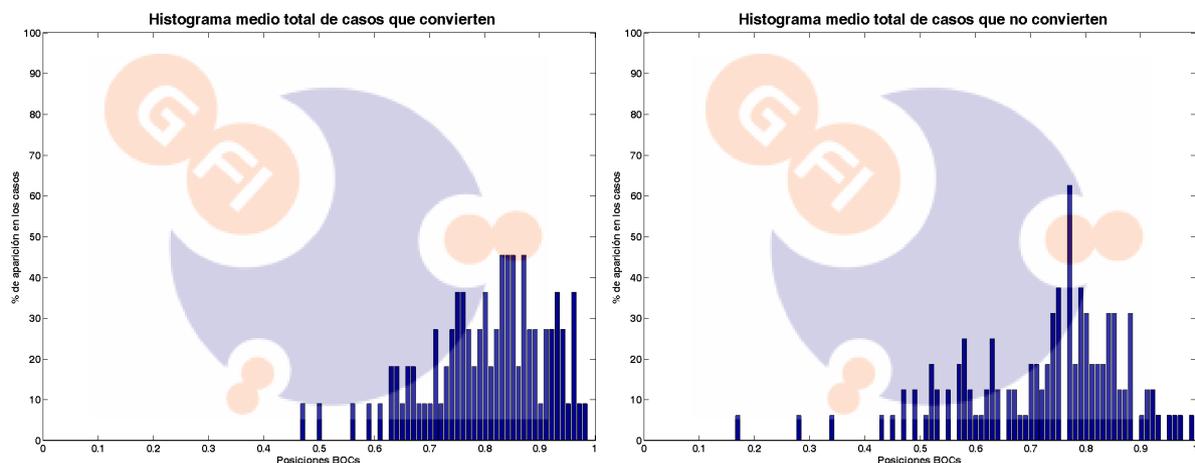
$$p \geq q_3 + w(q_3 - q_1) \quad (2.1)$$

$$p \leq q_1 + w(q_3 - q_1) \quad (2.2)$$

2.2. Histogramas

Los histogramas son representaciones gráficas ordenadas del número de concurrencias de un determinado elemento dentro de un conjunto de datos, que nos dan una visión general de la situación de los mismos.

En nuestro caso, al representar en la Figura 2.2 los casos que convierten y los que no convierten mediante este tipo de representación gráfica, podemos apreciar las claras diferencias entre estos dos conjuntos de datos.



(a) Representación de histograma de los casos que convierten. (b) Representación de histograma de los casos que no convierten.

Figura 2.2: Histogramas de los casos.

Así, vemos como para los casos que no convierten, el histograma nos muestra una mayor incidencia en la posición normalizada $x = 0,77$ y una mayor variabilidad, como ya habíamos podido ver en el diagrama de cajas de la Figura 2.1; no obstante los casos que convierten se muestran mucho más compactos alrededor de la posición normalizada $x = 0,8$ y no tienen una tendencia clara en sus bandas, con algún outlier por debajo de la posición $x = 0,6$.