

# Análisis Descriptivo PD-L1

Para este analisis la variable de interes es PD-L1 y su relacion con las demas variables variables.

Tabla 1: Frecuencia de la variable PD-L1

PD-L1	n
0	121
1	18
NA	191

Donde:

- **0 (Negativo)**: No se detectó PD-L1 en las células cancerosas.
- **1 (Positivo)**: Se detectó PD-L1 en las células cancerosas.
- **NA** : Sin resultado valido.

De acuerdo Tabla 1 en la variable **PDL1** encontramos 191 valores faltantes (NA), por tanto al eliminar esos valores faltantes nos quedamos con 139 observaciones.

Dado la naturaleza de la variables de intereres (binaria) el obketivo seria por medio de la regresion logistica lograr estimar la relación de una o más variables independientes con la variable dependiente que en este caso es PD-L1 y factores determinate en dicha variable.

Teniendo en cuenta el la tipologia de las variables independientes donde la mayoría son de tipo (categórico o factor), por tanto para la posible relacion que pueda haber entre la variable de interes (PD-L1) y las categoricas utilizaremos el El **test**  $\chi^2$  de independencia, también conocido como prueba  $\chi^2$  de Pearson, se emplea para estudiar si existe asociación entre dos variables categóricas y el **COEFICIENTE V DE CRAMER** funciona como una medida de relación estadística basada en chi cuadrado, se pueda precisar la fuerza de asociación entre dos o más variables. En este sentido, el resultado del coeficiente varía entre cero y uno (siendo cero un valor nulo de asociación)

## Visualización con las variables numericas

Primeramente observaremos la relación e las variables de tipo numérico con respecto a la variable de interes(PD-L1).

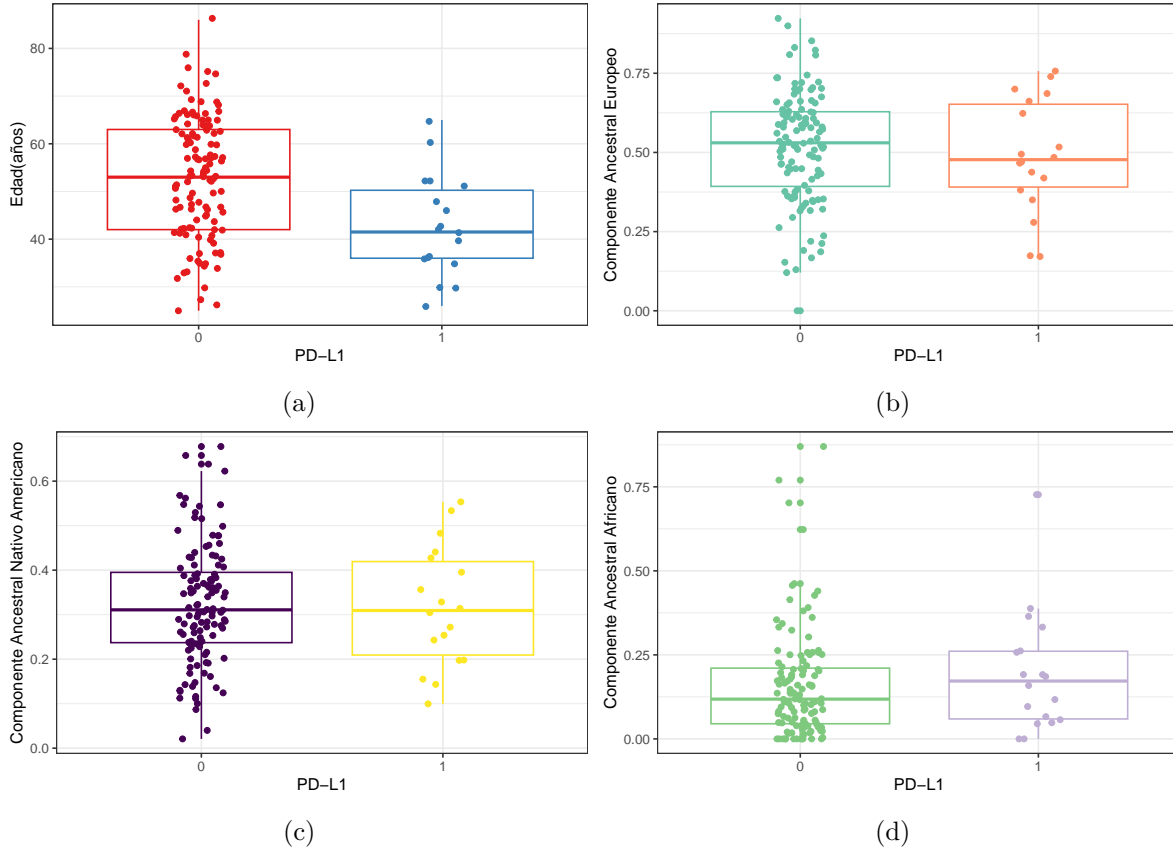


Figura 1: Diagrama de cajas

De la Figura 1 podemos concluir que:

- En la Figura 1a observamos hay mayor variación en la edad de las mujeres con PD-L1 negativo, las mujeres con PD-L1 positivo tienden más jóvenes en comparación con aquellas con PD-L1 negativo.
- En la Figura 1b observamos que las mujeres con PD-L1 positivo presentan un mayor porcentaje de ascendencia europea.
- En la Figura 1c observamos que existe una ligera mayor proporción de ascendencia nativo americana en el grupo PD-L1 positivo, mientras que el grupo negativo hay mayor variabilidad con el porcentaje de ascendencia nativo americana.

- En la Figura 1c observamos una mayor dispersión en el grupo PD-L1 negativo en la ascendencia africana, para los del grupo PD-L1 positivo no hay mucha variabilidad con respecto a la ascendencia africana.

Ademas en cada uno de los boxplots se observan indicios de que no hay diferencia significativas entre si es positivo o negativo con respecto a cada una de las variables, esta hipótesis sera verificada por la prueba de homogeneidad por medio del test chi-cuadrado donde la hipótesis nula es que las distribuciones de las dos poblaciones son iguales.

Tabla 2: p-values del test chi-cuadrado

Variables	p-valor
Edad-PDL1	0.3526
Eur-PDL1	0.46
Nam-PDL1	0.46
Afr-PDL1	0.5114

De acuerdo a la Tabla 2 podemos afirmar que no existe diferenecia entre los positivos y los negativos respecto a cada una de las variables.

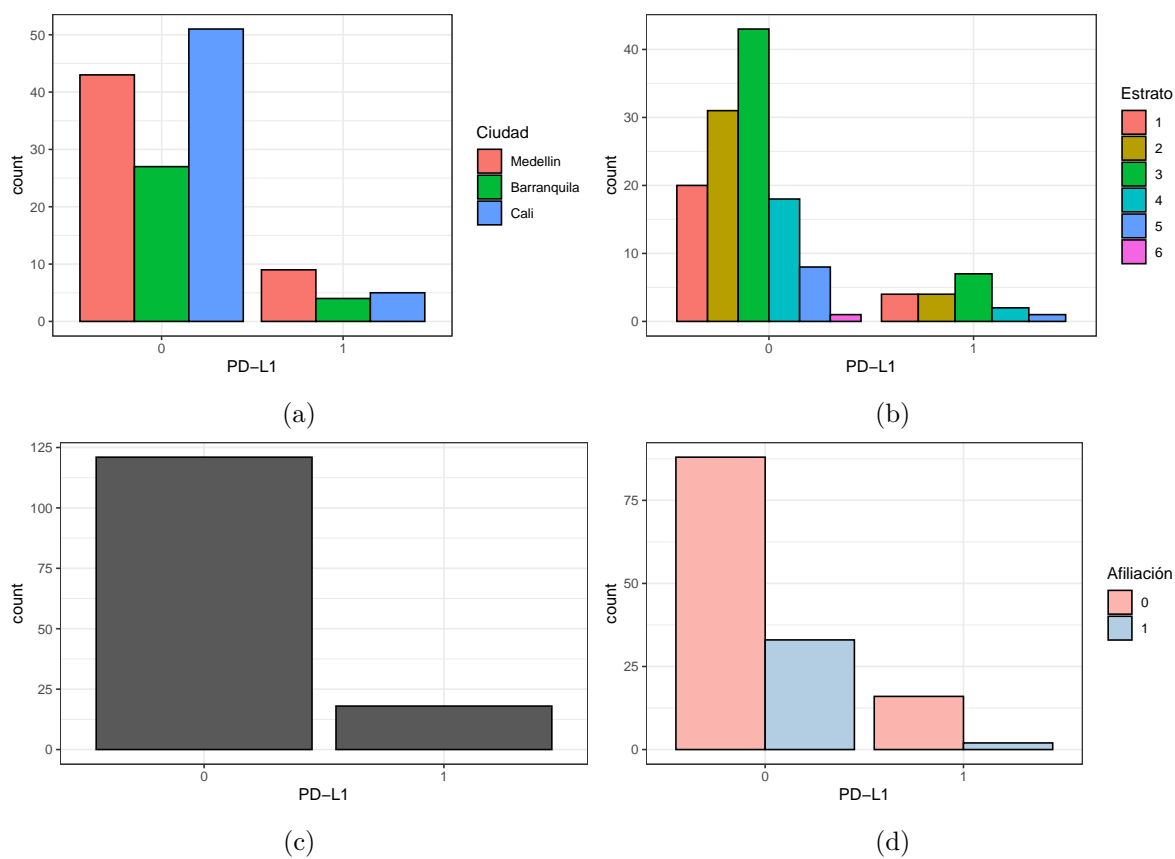


Figura 2: Graficos de Barras