

# Análisis Descriptivo PD-L1

Para este análisis la variable de interés es PD-L1 y su relación con las demás 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 con la tabla Tabla 1, la variable PD-L1 presenta 191 valores faltantes (NA). Al eliminar estos valores, el número de observaciones se reduce a 139. Es importante considerar que los valores faltantes representan más de la mitad de los registros, lo que puede generar sesgos en el análisis y afectar la capacidad de encontrar resultados significativos.

Teniendo en cuenta la tipología de las variables independientes, donde la mayoría son de tipo categórico, se utilizará la prueba de independencia mediante el **Test exacto de Fisher** para evaluar la posible relación entre la variable de interés (PD-L1) y las variables categóricas. Esta prueba permite determinar si existe una asociación significativa entre dos variables categóricas.

Además, para medir el tamaño del efecto, se empleará el **Coefficiente V de Cramer**, el cual indica la fuerza de asociación entre las variables. Los límites utilizados para su clasificación son los siguientes:

- Resultado entre 0 y 0,2 indica que **no hay asociación**
- Resultado de 0,2 indica una **asociación débil**
- Resultado entre 0,2 y 0,6 indica una **asociación moderada**
- Resultado entre 0,6 y 1 indica una **asociación fuerte**.

## Análisis con las variables numéricas

En primer lugar, analizaremos la relación entre las variables numéricas y la variable de interés (PD-L1) para identificar posibles asociaciones significativas.

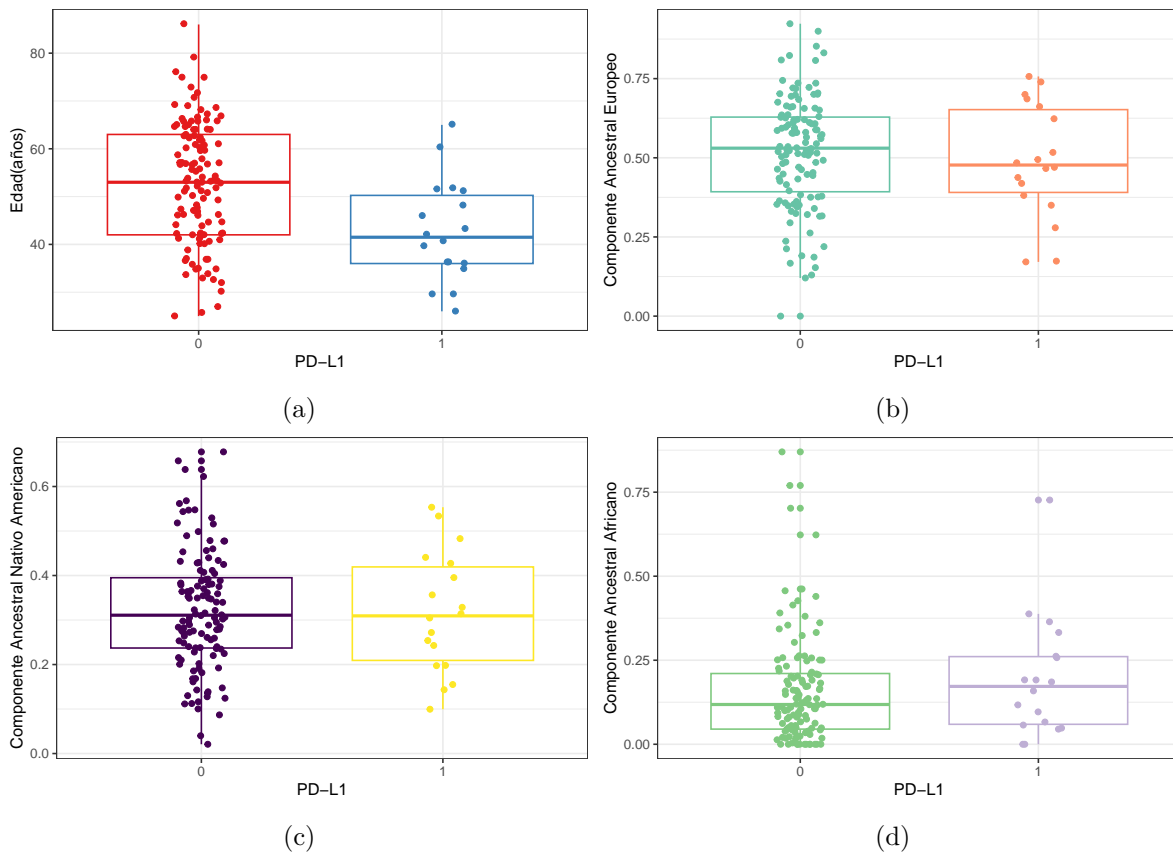


Figura 1: Diagrama de cajas

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.

Además, en los boxplots se observan indicios de que no existen diferencias significativas entre los grupos positivo y negativo con respecto a cada una de las variables analizadas. Esta hipótesis será verificada mediante una prueba de homogeneidad utilizando el test chi-cuadrado, donde la hipótesis nula establece que las distribuciones de ambas 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 con la tabla Tabla 2, podemos afirmar que no existen diferencias significativas entre los grupos positivo y negativo con respecto a cada una de las variables analizadas.

## Análisis con las variables categóricas

A continuación, analizaremos la relación entre la variable PD-L1 y las demás variables categóricas mediante el **Test exacto de Fisher**. En caso de encontrar una asociación significativa, se medirá el tamaño del efecto utilizando el **coeficiente V de Cramer**.

Mediante la prueba de independencia de Fisher, identificaremos las variables que presentan una asociación significativa con la variable PD-L1.

Tabla 3: P-values Test exacto de Fisher

Variables	P-value	Coefficiente V de Cramer
PDL1 - Grado Histológico	0.01253	0.2764
PDL1 - Er	4.156e-05	0.3662
PDL1 - Pr	0.0004695	0.3080
PDL1- Subtipo Molecular	0.0007031	0.3188

De la Tabla 3 podemos concluir que aunque la asociación es significativa, el tamaño del efecto (V de Cramer) indica que estas relaciones son de debil a moderada, por lo que PD-L1 puede estar influenciado por otros factores adicionales, la mayor asociación se observa con Er (0.3662), lo que podría indicar un papel relevante de PD-L1 en tumores con expresión de receptores hormonales.

## Conclusiones

La variable PD-L1 presenta un alto número de valores faltantes (191 de 330 registros, aproximadamente el 58%). Esto reduce significativamente el tamaño muestral a 139 observaciones y puede generar sesgos en las conclusiones obtenidas. No se encontraron diferencias significativas entre los grupos PD-L1 positivo y negativo en relación con las variables numéricas analizadas (edad y componentes ancestrales), la prueba de chi-cuadrado confirmó que las distribuciones de estas variables son similares en ambos grupos (p-valores mayores a 0.35 en todos los casos), sin embargo, se observa que las mujeres con PD-L1 positivo tienden a ser más jóvenes y a tener una mayor proporción de ascendencia europea.

Se identificó una asociación estadísticamente significativa entre PD-L1 y ciertas variables clínicas (grado histológico, expresión de receptores hormonales ER y PR, y subtipo molecular). El coeficiente V de Cramer indica que estas asociaciones son de débil a moderada, con la mayor asociación observada en la expresión de ER ( $V = 0.3662$ ), estos resultados sugieren que PD-L1 podría estar relacionado con la biología de los tumores que expresan receptores hormonales.