

# IDENTIFICACIÓN DE CARACTERÍSTICAS DE CÉLULAS DE CÁNCER DE MAMA POR MEDIO DE TESTORES TÍPICOS

---

VARIABLE: CONCAVIDAD (MEDIA)

ANALIZA LAS IRREGULARIDADES DE FORMA EN EL NÚCLEO DE LA CÉLULA. MIDEN EL NÚMERO Y LA SEVERIDAD DE LAS CONCAVIDADES Y HENDIDURAS EN EL NÚCLEO DE LA CÉLULA.

YULEIDIS MESA

# ESTADÍSTICA DESCRIPTIVA

## ▼ Células benignas

```
✓ [51] cancer_b["Concavity (mean)"].describe()  
0s
```

```
count    357.000000  
mean      0.046058  
std       0.043442  
min       0.000000  
25%      0.020310  
50%      0.037090  
75%      0.059990  
max       0.410800  
Name: Concavity (mean), dtype: float64
```

## ▼ Células malignas

```
✓ [52] cancer_m["Concavity (mean)"].describe()  
0s
```

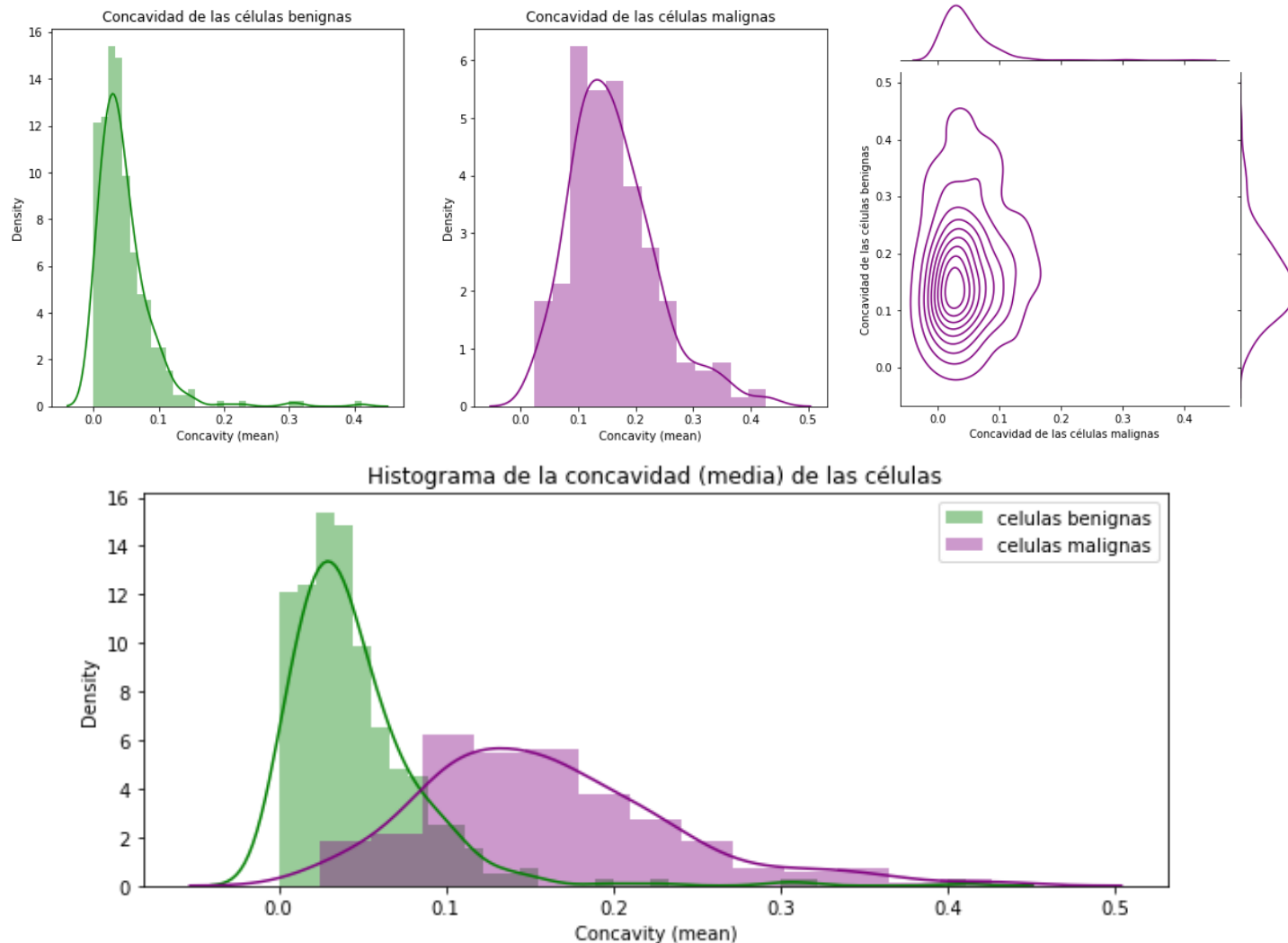
```
count    212.000000  
mean      0.160775  
std       0.075019  
min       0.023980  
25%      0.109525  
50%      0.151350  
75%      0.203050  
max       0.426800  
Name: Concavity (mean), dtype: float64
```

La mediana de la concavidad de las células del cáncer benigno es 0.037

La mediana de la concavidad de las células del cáncer maligno es 0.151

Se encontraron un total de 357 y 212 observaciones para las células benignas y malignas respectivamente. En términos numéricos se observa que la media de las células malignas es ligeramente mayor y se ubica en 0,16, con valores mínimos y máximos entre (0,024 y 0,427). Por su parte la desviación estándar se ubica para las células benignas en 0,04 y para las malignas en 0,075 por lo que podría suponer mas dispersión entre los datos para las segundas. Finalmente la mediana expone que el 50% de los datos se entra ubicado para las células benignas en un 0,037 y en 0,151 para las células malignas.

# DISTRIBUCIONES



Los histogramas reflejan distribuciones diferentes para la misma variable en los dos tipos de células, para el caso de las benignas se observa que la curva está más sesgada a la izquierda y los datos se encuentran mas agrupados. Mientras que para el caso de las células malignas los datos se distribuyen por todo el eje x, lo que da una impresión de distribución normal.

Por su parte, al superponer ambas distribuciones, se evidencia que la mediana de la distribución de las células malignas se encuentra más a la derecha, por lo se podría inferir sin evidencia estadística que es mayor.

# TEST

```
check_normality(concavity_b)
check_normality(concavity_m)
check_variance_homogeneity(concavity_b, concavity_m)
```

```
p value:0.0000
Reject null hypothesis >> The data is not normally distributed
p value:0.0000
Reject null hypothesis >> The data is not normally distributed
p value:0.0000
Reject null hypothesis >> The variances of the samples are different.
```

Se comprobó que los conjuntos de datos no cumplen los criterios necesarios para realizar un test paramétrico, no se cumplió la normalidad de la distribución, ni la homogeneidad de las varianzas. Se procedió a realizar el test no paramétrico Mann Whitney U, planteado la siguiente hipótesis:

$H_0: (\mu_m < \mu_b)$  La media de la concavidad para células malignas es menor que la media de la concavidad para las células benignas.

$H_1: (\mu_m > \mu_b)$  La media de la concavidad para células malignas es mayor que la media de la concavidad para las células benignas.

```
ttest,pvalue = sc.stats.mannwhitneyu(concavity_b, concavity_m, alternative="two-sided")
print("p-value:%.4f" % pvalue)
if pvalue < 0.05:
    print("Reject null hypothesis")
else:
    print("Fail to reject null hypothesis")
```

```
p-value:0.0000
Reject null hypothesis
```

Dándonos como resultado: Se rechaza la hipótesis nula, es decir que podemos decir con un nivel de confianza del 95% que la media de la concavidad para células malignas es mayor que la media de la concavidad para las células benignas.