

# Curvas ROC y estimación área bajo la curva.

---

MTRA. MELODY TREVIÑO RODRIGUEZ



# Curvas ROC

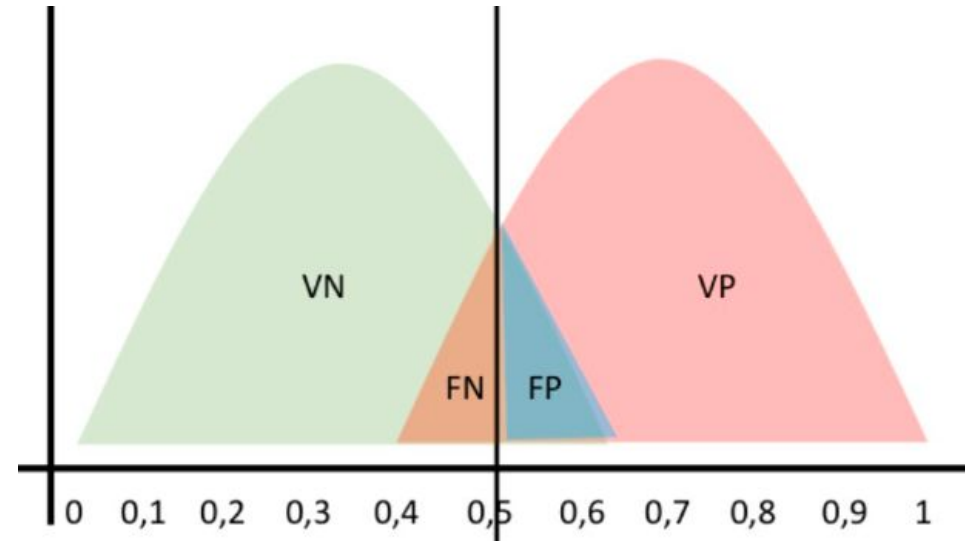
La curva ROC (Receiver Operating Characteristic o Característica Operativa del Receptor) es un método para evaluar clasificadores mediante una representación gráfica del rendimiento del clasificador que muestra la distribución de las fracciones de verdaderos positivos y de falsos positivos.

La fracción de verdaderos positivos se conoce como **sensibilidad**, sería la probabilidad de clasificar correctamente a un individuo cuyo estado real sea definido como positivo.

La **especificidad** es la probabilidad de clasificar correctamente a un individuo cuyo estado real sea clasificado como negativo.

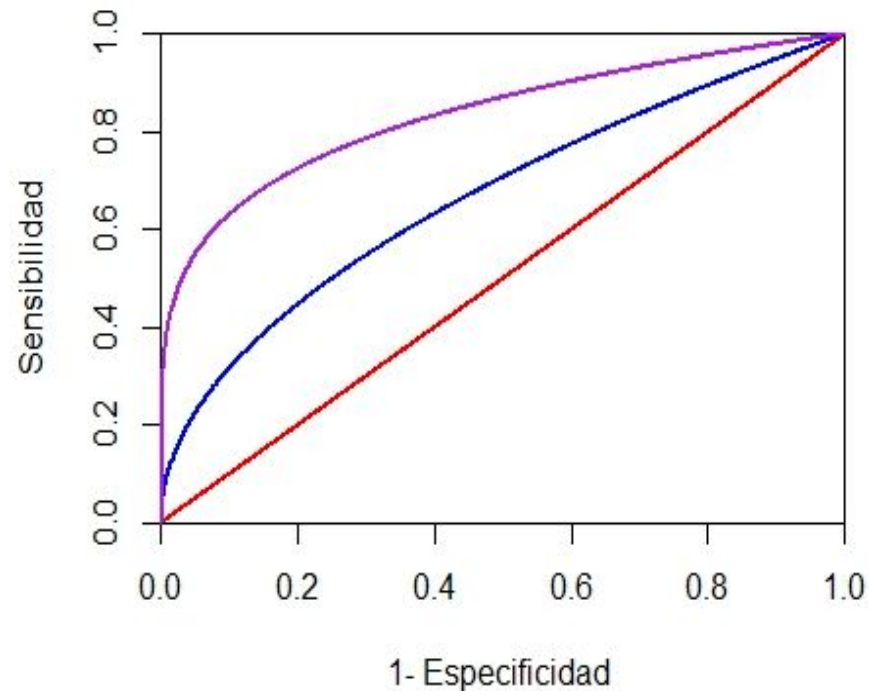
La curva ROC es conocida como la representación de **sensibilidad vs (1-especificidad)**

$$1 - \text{especificidad} = \frac{FP}{VN + FP}$$



Tipos de errores		Realidad	
		P = Positivos	N = Negativos
Predicción	P = Positivos	VP	FP (Error tipo I)
	N = Negativos	FN (Error tipo II)	VN

# Curvas ROC



Cada resultado de predicción representa un punto en el espacio ROC. Es decir, cada punto de la curva ROC corresponde a un posible punto de corte del test diagnóstico, y nos informa su respectiva sensibilidad (eje Y) y 1-especificidad (eje X). La línea trazada desde el punto 0,0 al punto 1,1 recibe el nombre de diagonal de referencia, o línea de no-discriminación.

El mejor método posible de predicción se situaría en un punto en la esquina superior izquierda, o coordenada (0,1) del espacio ROC, representando un 100% de sensibilidad (ningún falso negativo) y un 100% también de especificidad (ningún falso positivo).

Una clasificación totalmente aleatoria daría un punto a lo largo de la línea diagonal, que se llama también línea de no-discriminación, es decir, un modelo inútil.

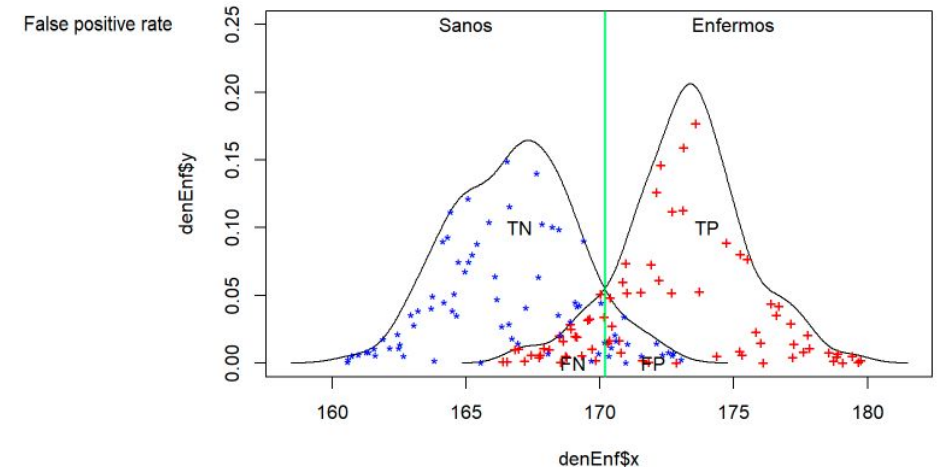
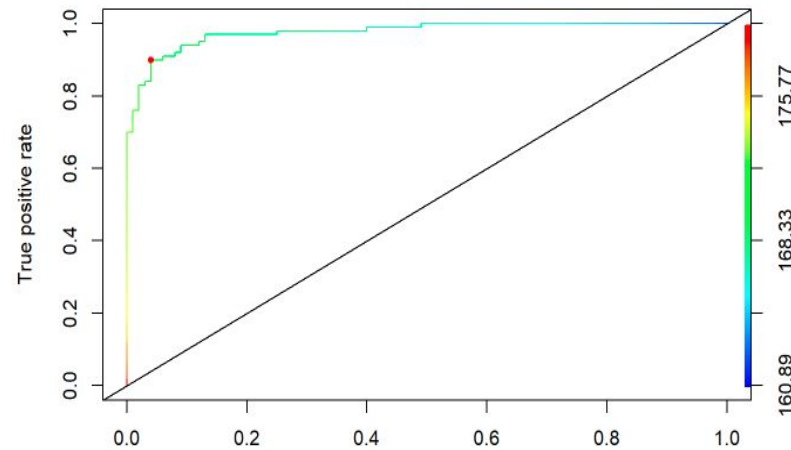
Por lo tanto, lo que se espera encontrar son puntos sobre una curva por encima de la diagonal y entre más cercana se encuentre de la esquina superior izquierda, mejor será su predicción.

# Curvas ROC

El **punto de corte óptimo** es aquel que maximiza el cociente de verdaderos positivos (sensibilidad), y que al mismo tiempo minimiza el ratio de falsos positivos (1-especificidad).

Desde una perspectiva gráfica, será el punto de la curva que quede lo más arriba y a la izquierda posible.

Es llegar a un compromiso de maximización de ambos parámetros, teniendo en cuenta que aumentar uno significa disminuir el otro.

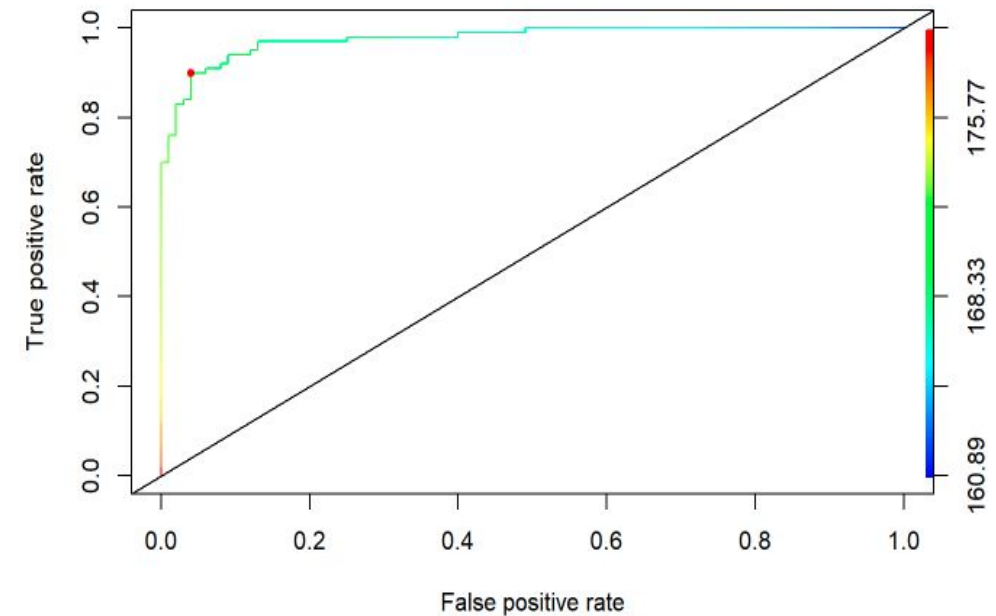


# Estimación área bajo la curva

El valor del área que queda bajo la curva ROC (AUC), puede interpretarse como la probabilidad de que, ante dos individuos, uno de una distribución y otro de otra, la prueba clasifique a los dos correctamente.

El AUC se sitúa siempre entre 0 y 1, cuanto mayor sea el valor del AUC mejor será la prueba diagnóstica.

Un valor cercano al 0.5 (50% del área total) será un valor bastante malo. dependiente del problema de investigación, se podría ser más o menos exigente con el clasificador.



# Estimación área bajo la curva (AUC)

Para calcular el AUC, se integra esta función  $TP(FP)$  con respecto a  $FP$  utilizando el método de integración numérica.

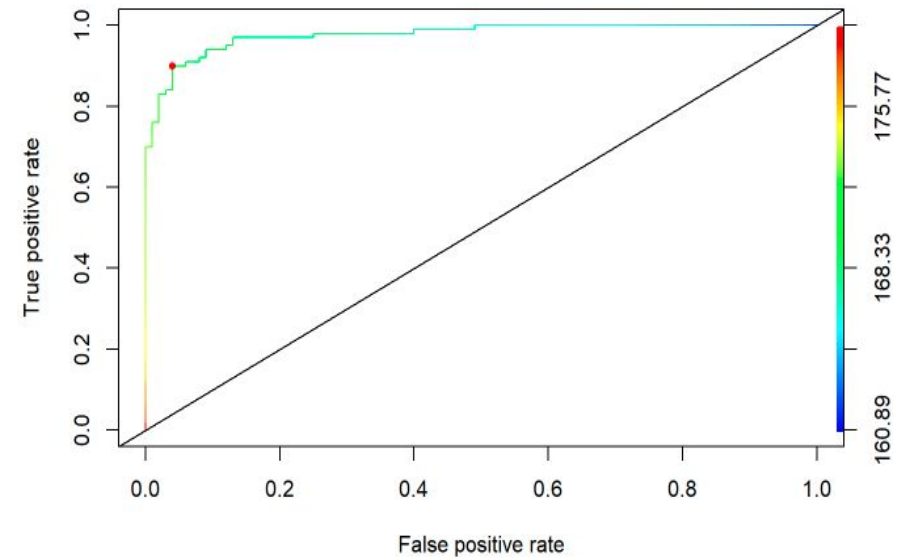
$$AUC = \int_0^1 TP(FP) dFP$$

La integral se evalúa desde 0 hasta 1 porque la tasa de falsos positivos varia en ese rango.

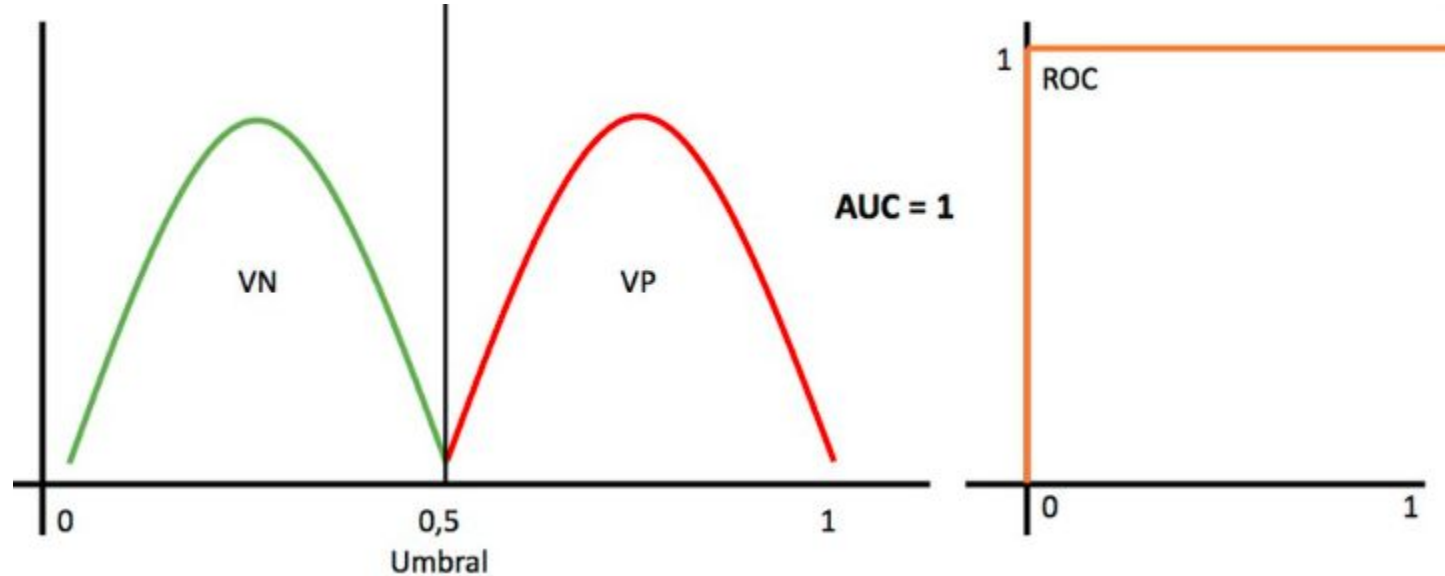
Si  $FP_i$ ,  $TP_i$ , son puntos discretos se puede aproximar la integral utilizando métodos numéricos como la regla del trapecio;

$$AUC \approx \sum_{i=1}^{n-1} \frac{(TP_{i+1} + TP_i) \cdot (FP_{i+1} - FP_i)}{2}$$

Donde  $n$  es el número total de puntos de la curva ROC



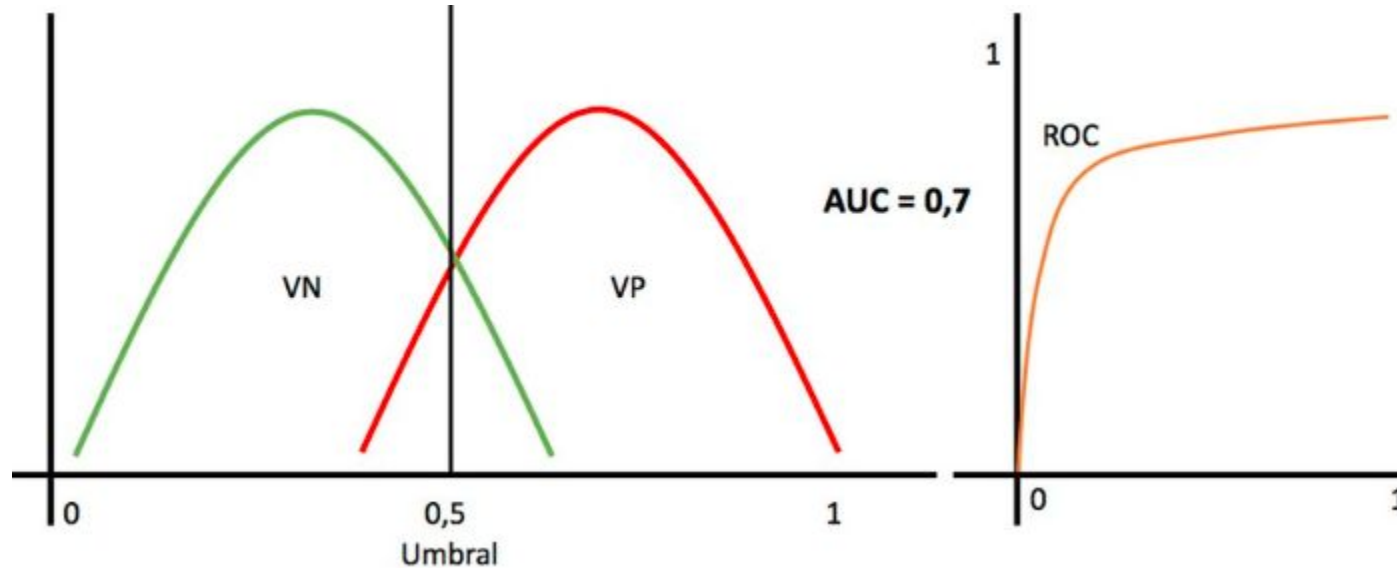
# Estimación área bajo la curva



Situación ideal. Cuando dos curvas no se superponen en absoluto, el modelo tiene una medida ideal de separación. Es perfectamente capaz de distinguir entre clase positiva y clase negativa.

Área bajo la curva (AUC)=1

# Estimación área bajo la curva

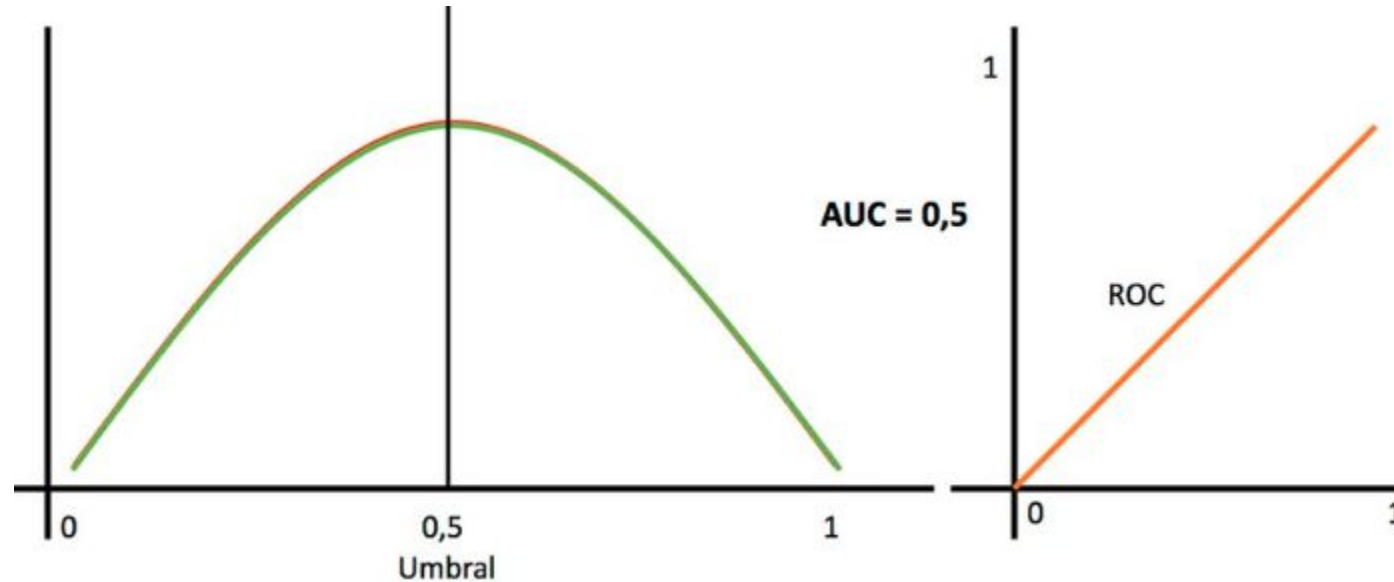


Cuando AUC es 0.7, significa que hay 70% de probabilidad de que el modelo pueda distinguir entre clase positiva y clase negativa.



# Estimación área bajo la curva

---



Cuando el AUC es aproximadamente 0.5, el modelo no tiene capacidad de discriminación para distinguir entre clase positiva y clase negativa.

## Tarea sobre **Comparación de curvas ROC en muestras**

---

1) Investigar y describir por lo menos 1 ejercicio donde apliquen las curvas ROC, puede ser aplicar las curvas ROC a ejercicio de clasificaciones previas que han realizado en semestres anteriores:

- Realizan una parte introductoria que describa el problema de investigación, los resultados obtenidos etc.
- La aplicación del método de la curva ROC
- Describir lo que se hace en cada paso y aportar las conclusiones personales a las que llegaste con el resultado obtenido, se expondrá la siguiente clase