

# Contents

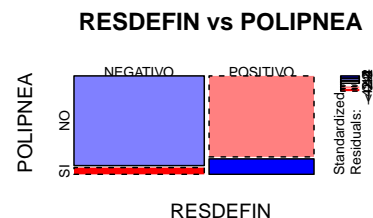
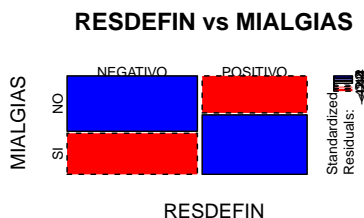
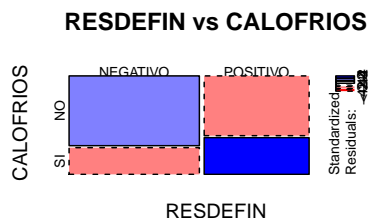
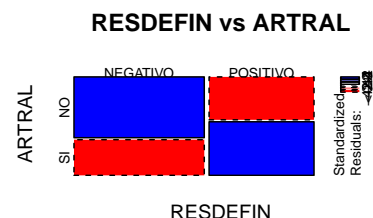
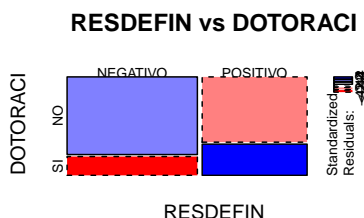
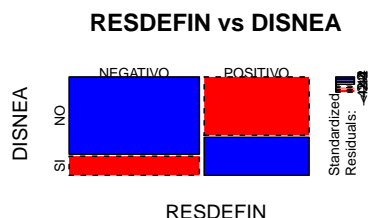
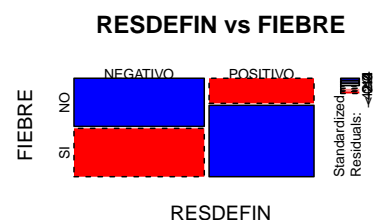
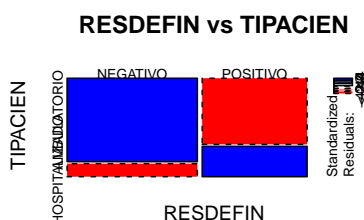
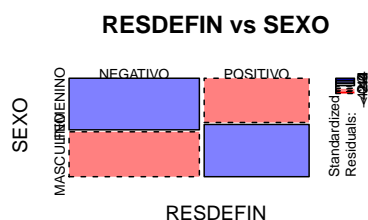
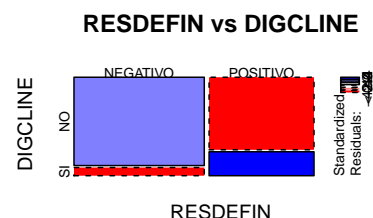
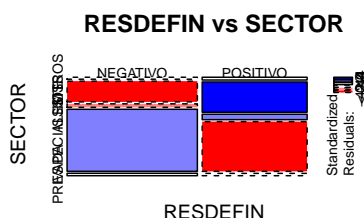
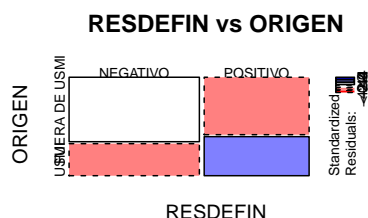
<b>Manipulación de los datos</b>	<b>2</b>
<b>RESDEFIN</b>	<b>2</b>
Positivos a Covid . . . . .	3
<b>Evolución</b>	<b>4</b>
Descriptivo . . . . .	5
<b>Resultados</b>	<b>7</b>
<b>Conclusión</b>	<b>8</b>

# Manipulación de los datos

Para el análisis que se llevará a cabo en este proyecto, además del preprocesamiento que se hizo en clase sobre las variables de ocupación, sector y resultado definitivo se imputarán los “SE IGNORA” de la mayoría de variables dicotómicas de manera aleatoria con probabilidad justa, además creamos la variable que mide los días transcurridos entre el inicio de síntomas y el ingreso al sistema.

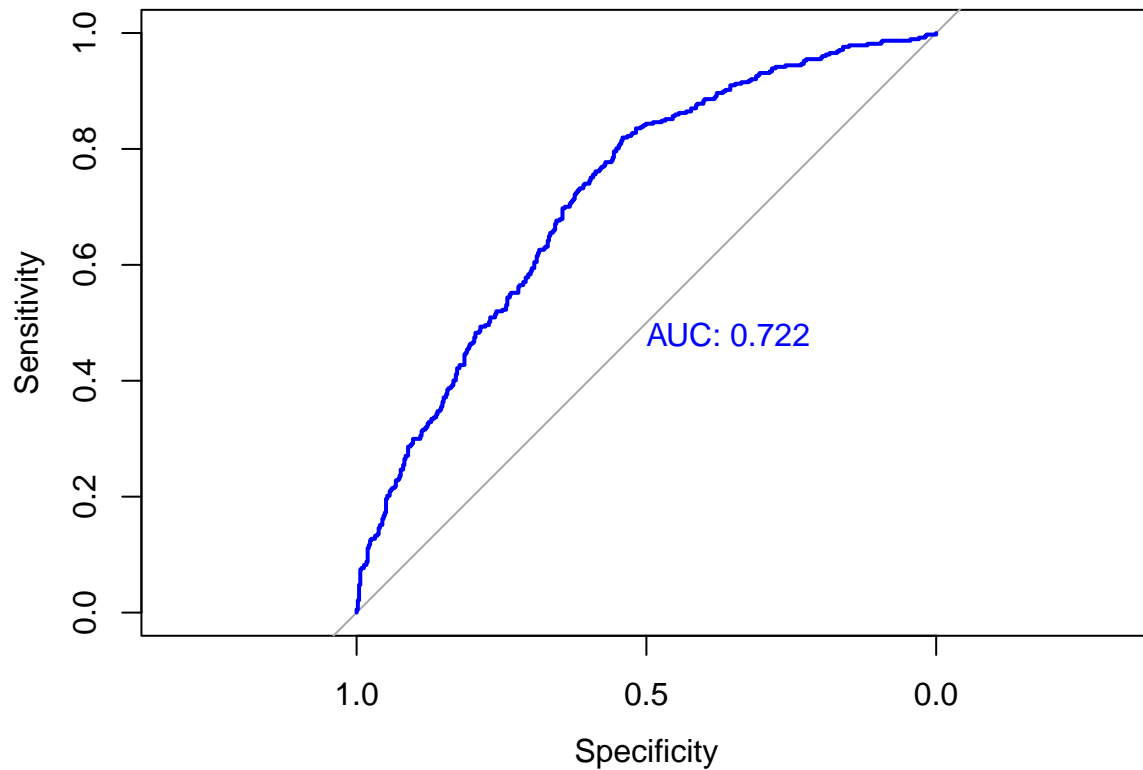
## RESDEFIN

##Descriptivo





mostrarán posteriormente, pero se puede observar la curva ROC de sensibilidad/especificidad con el método de regresión logística



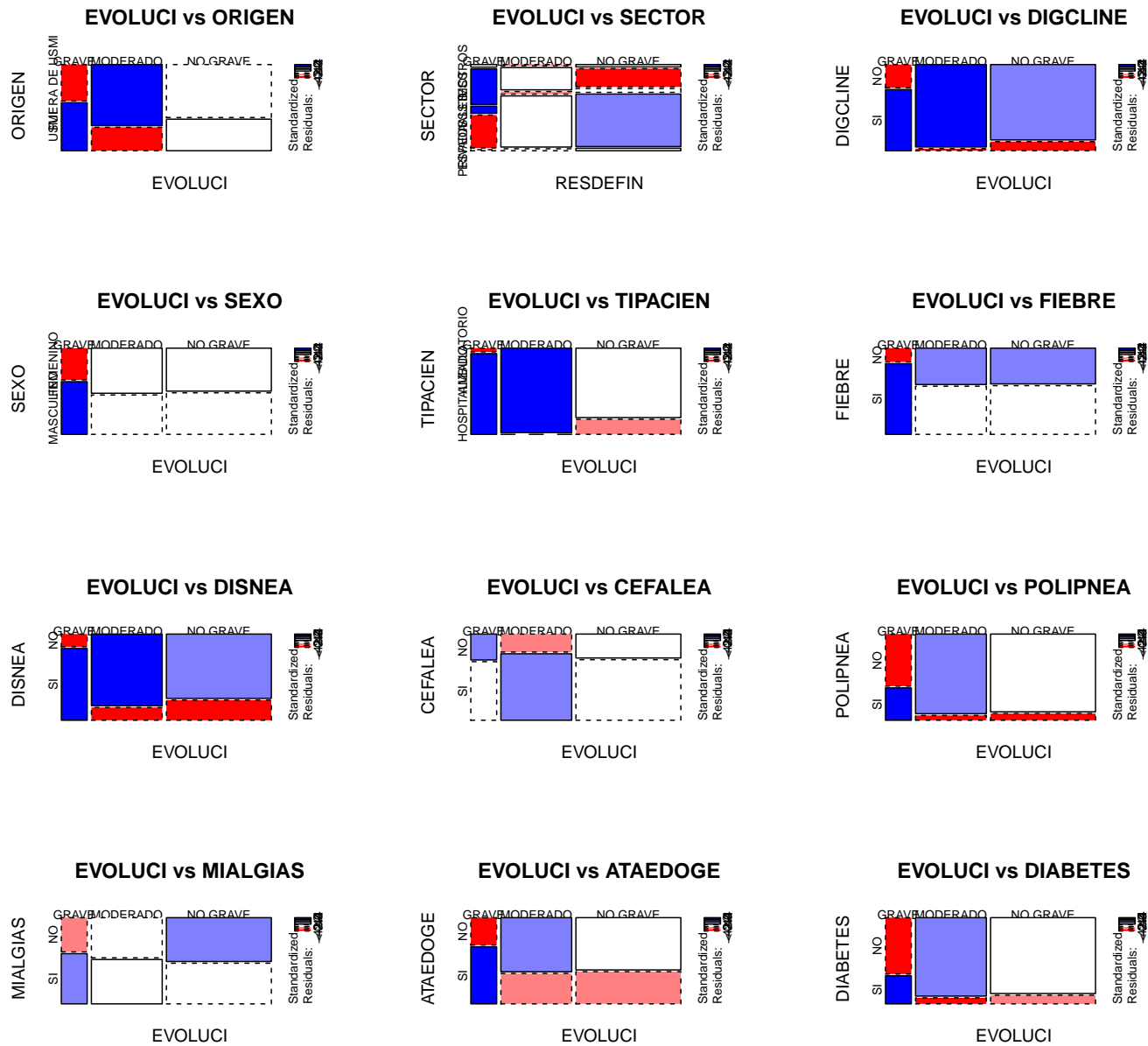
```
##
## Call:
## roc.default(response = df$RESDEFIN[-index], predictor = p, plot = TRUE,      print.auc = TRUE, col = 
##
## Data: p in 473 controls (df$RESDEFIN[-index] NEGATIVO) < 377 cases (df$RESDEFIN[-index] POSITIVO).
## Area under the curve: 0.7217
```

## Evolución

Por otro lado, también trataremos de explicar la variable evolución, la cual la dividiremos en 2 y 3 categorías, **GRAVE** y **NO GRAVE**, y **GRAVE, MODERADO** y **NO GRAVE** respectivamente.

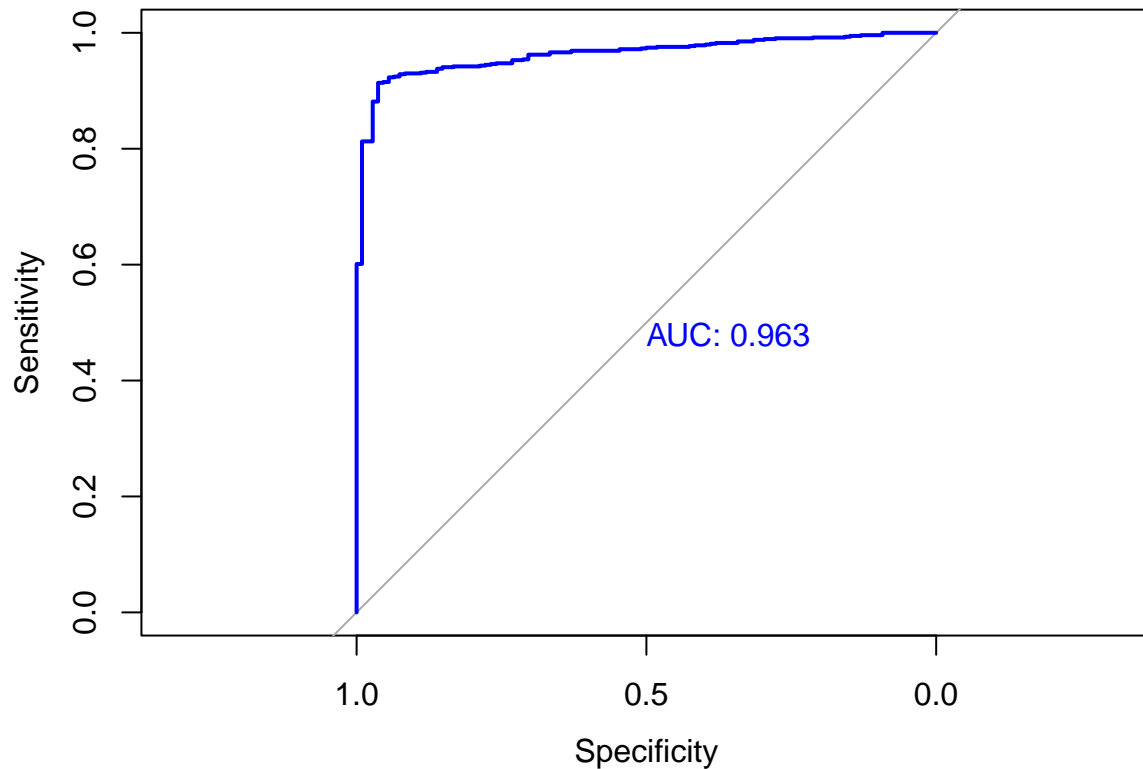
Las variables explicativas son las mismas que para el caso anterior además de las comorbilidades. Al usar sólo dos categorías haremos lo mismo que con la variable **RESDEFIN**, primero usar regresión logística y un árbol, seleccionar variables y posteriormente usar los cuatro métodos.

Descriptivo



##Metodología





```
##
## Call:
## roc.default(response = df$EVOLUCI[-index], predictor = p, plot = TRUE,      print.auc = TRUE, col = "blue")
##
## Data: p in 108 controls (df$EVOLUCI[-index] GRAVE) < 742 cases (df$EVOLUCI[-index] NO GRAVE).
## Area under the curve: 0.9634
```

Y al tener tres categorías usaremos las mismas variables seleccionadas anteriormente con los cuatro clasificadores

## Resultados

```
##          LOG/MULT      SVM  BOSQUES      NB
## RESDEFIN      0.6690224 0.6822069 0.6605470 0.6647838
## EVOLUCION BIN 0.9218441 0.9296127 0.9140771 0.9213732
## EVOLUCION TRI 0.6007425 0.6181709 0.5899832 0.5863895
```

Finalmente, habiendo analizado que en general los métodos se comportan de manera similar, mostraremos una matriz de confusión para cada problema de clasificación con diferentes métodos: logístico para resultado definitivo, SVM para evolución con dos categorías y NB para evolución con tres categorías.

```
##          Reference
## Prediction NEGATIVO POSITIVO
```

##	NEGATIVO	347	169
##	POSITIVO	126	208

##		Reference	
##	Prediction	GRAVE	NO GRAVE
##	GRAVE	79	35
##	NO GRAVE	29	707

##		Reference		
##	Prediction	GRAVE	MODERADO	NO GRAVE
##	GRAVE	95	1	45
##	MODERADO	0	166	166
##	NO GRAVE	13	139	225

Donde la matriz de confusión para resultado definitivo tiene una tasa de error importante en ambas clases, pero la mayoría sigue cayendo en donde supondríamos. En la de tres clases de evolución se confunden mucho entre moderados y graves. En la de dos clases las matrices de confusión parecen comportarse bien. Para todos los métodos las matrices de confusión tienen un comportamiento similar.

## Conclusión

De acuerdo a los resultados anteriores, podemos observar que tenemos una mejor precisión de clasificación de los métodos usados para el modelo Evolución con dos categorías que para el modelo con 3 categorías, esto puede deberse al desequilibrio en la baja cantidad de casos graves en comparación con los no graves, por lo que el acierto de clasificación aumentaría.

Debido a los buenos resultados al menos para un modelo, entonces hemos encontrado bastante útil la aplicación de los diferentes métodos de clasificación supervisada para un problema con una muestra relativamente grande y con datos reales.