



# ANDRES DE LA ROSA

## PROYECTO TITANIC



# OBJETIVOS

- Determinar las características de los pasajeros del Titanic que más influyeron a que estos no sobrevivieran al siniestro.



# BREVE ANÁLISIS DESCRIPTIVO

Este análisis se realizó con el objetivo de crear simulaciones para el modelo lo mas reales posibles y para la imputación de la edad.

## Análisis de género

- El 74.20% de las personas que sobrevivieron eran mujer.

## Análisis de las clases

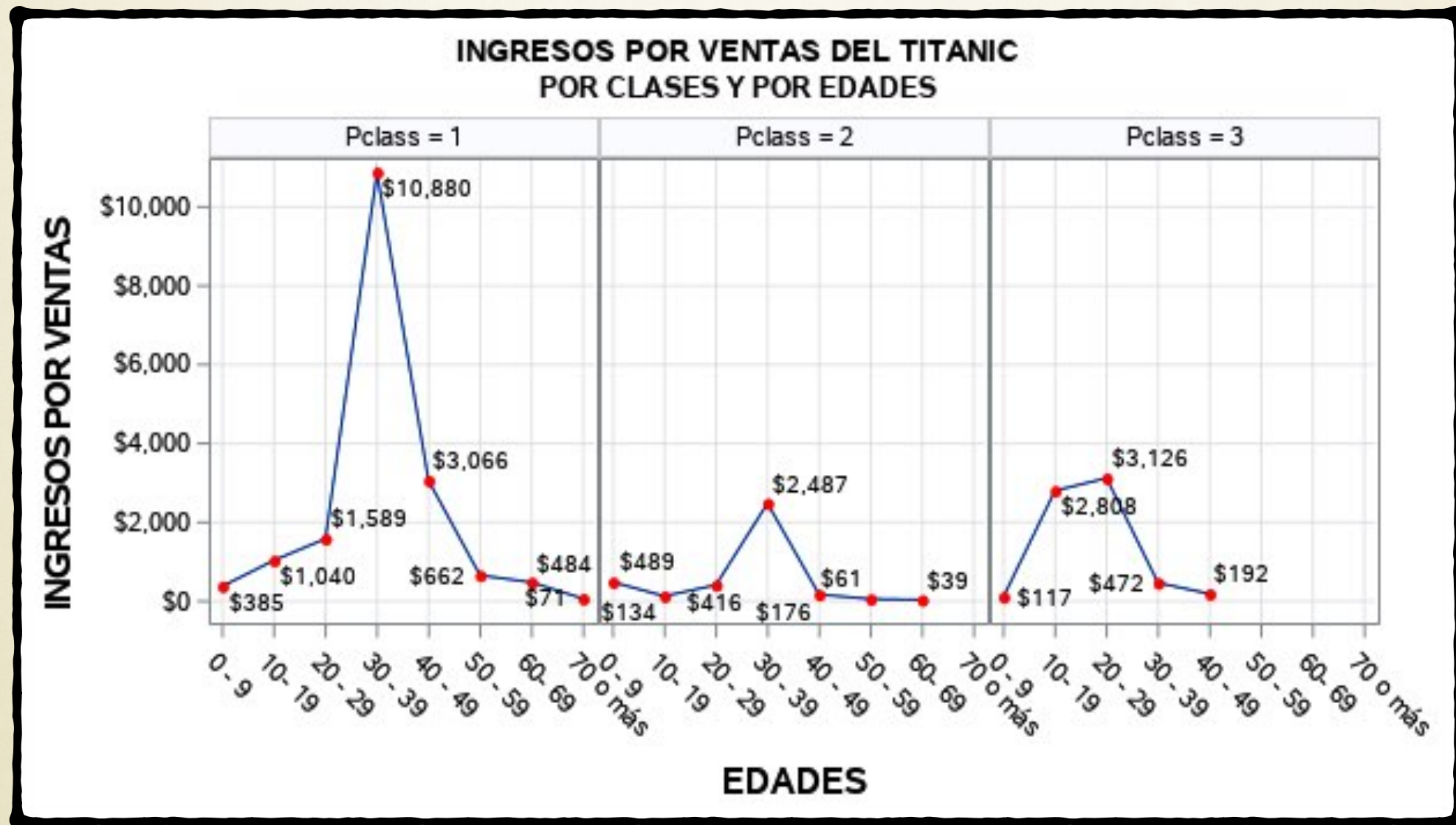
- **Los jóvenes se concentraron en la tercera clase.** La edad promedio de la tercera clase era 25.36 años, la edad promedio de la primera clase era 38 años y la edad promedio de la segunda clase era 30 años (bajo un ANOVA no balanceado se determinó que esta diferencia de edades es estadísticamente significativa al 5% nivel de significancia ( $P < 0.000$  &  $F = 188.14$ ) (Varianza no homogénea)
- El 67.76% de las personas que no sobrevivieron eran de **la tercera clase**, siendo esta **la clase que mas fatalidades sufrió**. El 39.77% de las personas que sobrevivieron eran de la primera clase, siendo esta la clase que menos fatalidades sufrió.

## Análisis de las edades

- **No existen diferencias** estadísticamente significativas **entre la edad promedio de los que sobrevivieron** (28.858821) **y los que no sobrevivieron** (29.952904) al 10% nivel de significancia (ver diagnósticos adjuntos).
- Las mujeres (n=314) tenían una edad promedio de 26.93 años y los hombres (n=577) tenían una edad promedio de 30.944, estas diferencias no son estadísticamente significativas al 5% nivel de significancia (no se cumplen las condiciones de normalidad). Así mismo no existen diferencias significativas en la edad entre hombres y mujeres que sobrevivieron y que no sobrevivieron, sabiendo que para ninguno de los casos se cumplen las suposiciones de normalidad.

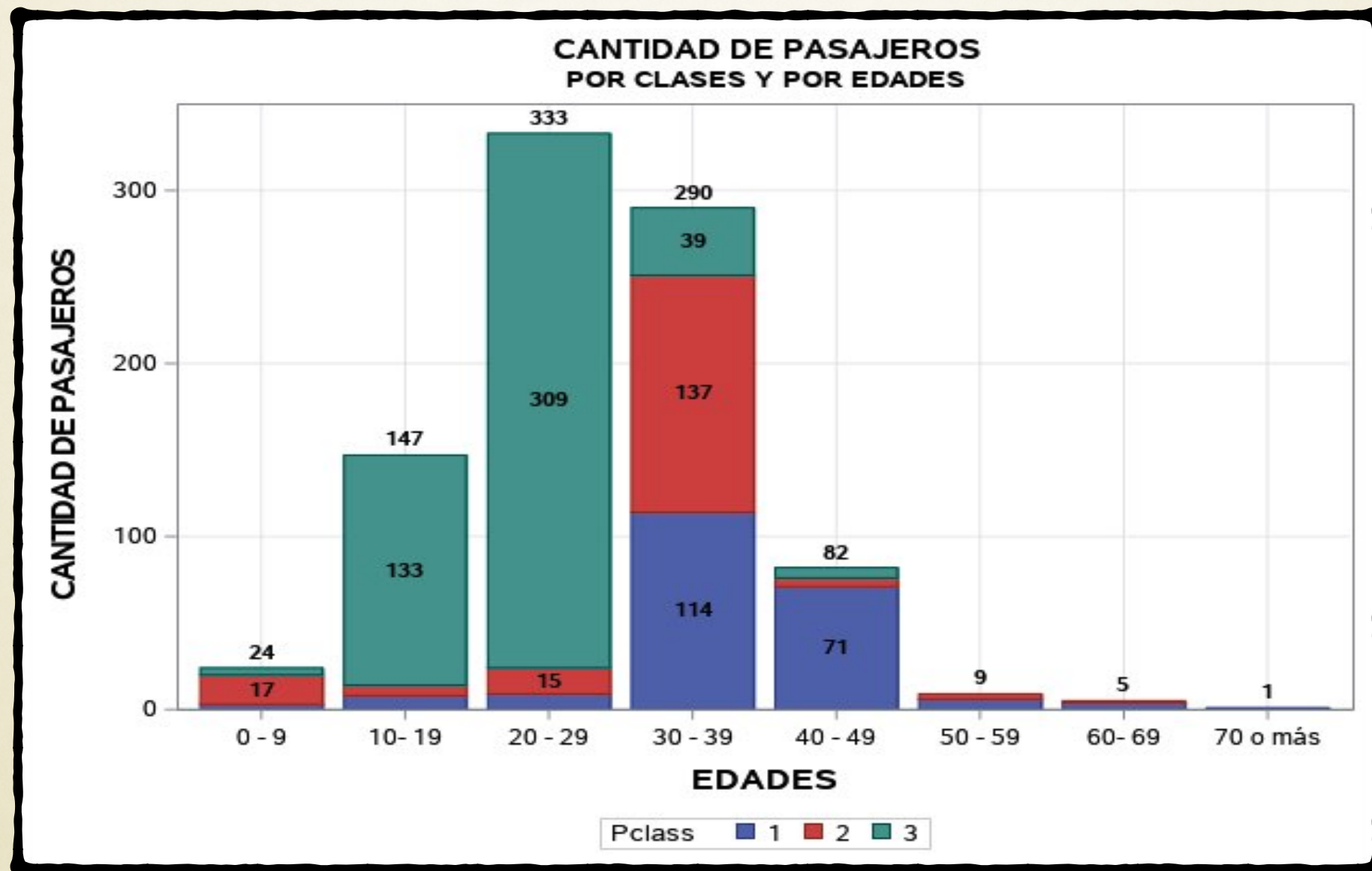


# BREVE ANÁLISIS DESCRIPTIVO



Los ingresos del Titanic, dadas las tarifas, para la primera clase se concentró en pasajeros entre 30-49 años, para la segunda clase en pasajeros que tenían entre 30-39 años y para la tercera clase en pasajeros que tenían entre 10 y 29 años.

# BREVE ANÁLISIS DESCRIPTIVO



Si vemos la cantidad de pasajeros por edad y por clase, vemos que la mayoría de las personas de la 3ra clase tienen entre 10 y 29 años. Este comportamiento va de la mano con el gráfico anterior.



# MODELO Y RESULTADOS

Referencias de las variables categóricas (Effect Coding)

**Clase:** Tercera clase

**Sexo:** Masculino

**Puerto de embarque:** Southampton, Inglaterra  
(de donde salen la mayoría de los pasajeros) los otros  
puertos fueron Cherbourg Francia y Queenstown, Irlanda

$$\text{Supervivencia} = \beta_1 \text{Clase} + \beta_2 \text{Sexo} + \beta_3 \text{PuertodeEmbarque} + \beta_4 \text{Edad} + \beta_5 \text{CantidaddeFamiliares} + \beta_6 \text{Tarifa}$$

## Matriz de Confusión

| Observado | Estimaciones |          |       |
|-----------|--------------|----------|-------|
|           | Positivo     | Negativo | Total |
| Positivo  | 243          | 99       | 342   |
| Negativo  | 76           | 473      | 549   |
| Total     | 572          | 319      | 891   |

**Hyperparameter:** cutoff= 0.50

Misclassification Rate: 19.64%

False Positive Rate: 17%

False Negative Rate: 24%

True Positive Rate (Sensitivity): 83%

True Negative Rate (Specificity): 76%



# MODELO Y RESULTADOS

## Parámetros

## Odds Ratio

| Analysis of Maximum Likelihood Estimates |        |    |          |                |                 |            |
|--|--------|----|----------|----------------|-----------------|------------|
| Parametro                                | Nivel  | DF | Estimate | Error Estándar | Wald Chi-Square | Pr > ChiSq |
| Intercepto                               |        | 1  | -0.8723  | 0.3692         | 5.5813          | 0.0182     |
| Clase                                    | 1      | 1  | 2.6988   | 0.3398         | 63.0823         | <.0001     |
| Clase                                    | 2      | 1  | 1.4972   | 0.2426         | 38.0748         | <.0001     |
| Sexo                                     | female | 1  | 2.4696   | 0.1987         | 154.5006        | <.0001     |
| Puerto de embarque                       | C      | 1  | 0.375    | 0.2399         | 2.4436          | 0.118      |
| Puerto de embarque                       | Q      | 1  | 0.4161   | 0.3424         | 1.4766          | 0.2243     |
| Edad                                     |        | 1  | -0.0545  | 0.012          | 20.5311         | <.0001     |
| Cantidad de familiares                   |        | 1  | -0.2767  | 0.13           | 4.5331          | 0.0332     |
| Tarifa                                   |        | 1  | 0.000406 | 0.00226        | 0.0322          | 0.8576     |

La significancia de los coeficientes logit nos sugieren que el coeficiente del puerto de embarque no es estadísticamente significativo al nivel de significancia del 10% para ninguno de los puertos comparado con el puerto de Southampton.

| Association of Predicted Probabilities and Observed Responses |         |           |       |
|---|---------|-----------|-------|
| Percent Concordant  | 85.4    | Somers' D | 0.711 |
| Percent Discordant  | 14.3    | Gamma     | 0.714 |
| Percent Tied  | 0.3     | Tau-a     | 0.338 |
| Pairs   | 181,365 | c         | 0.856 |

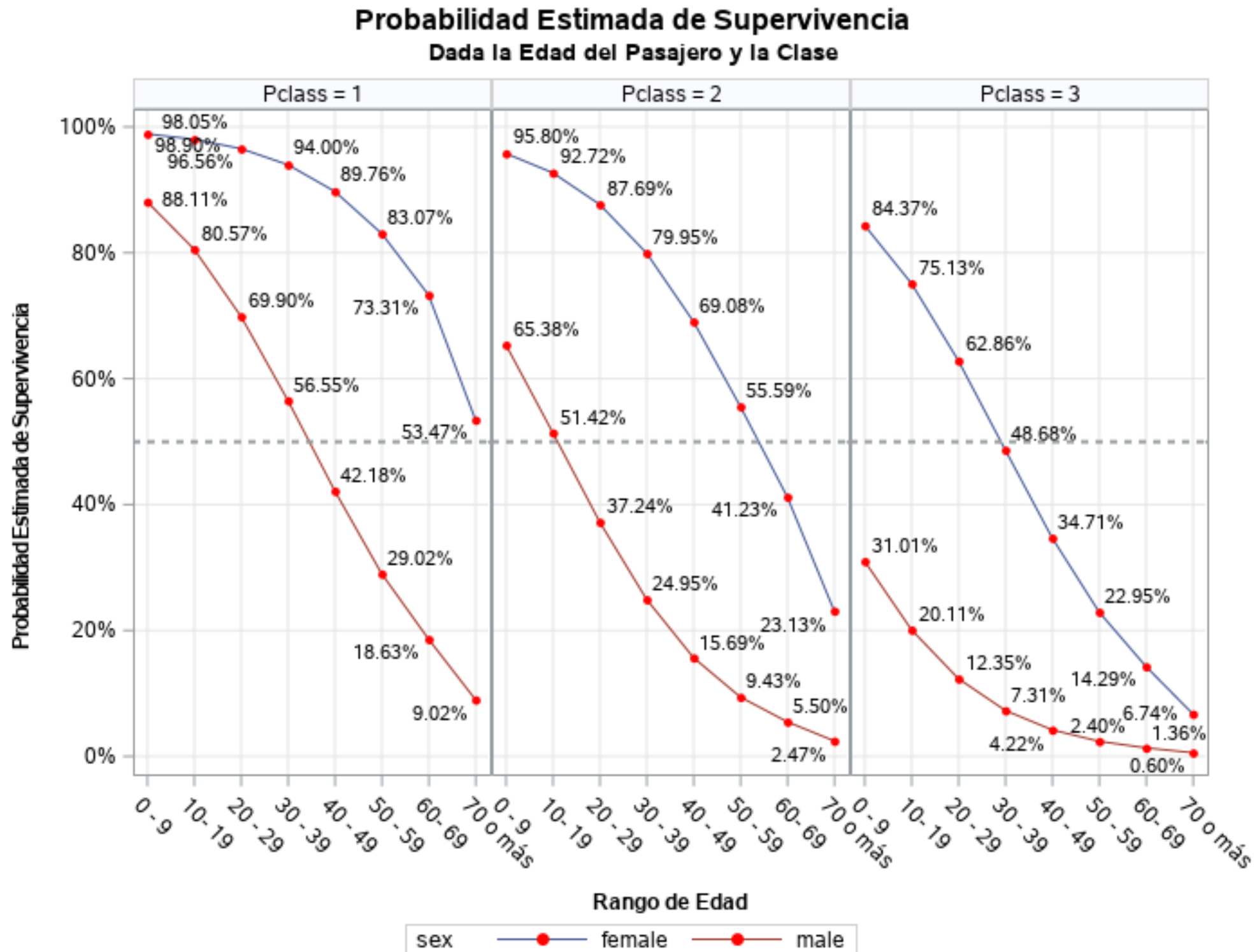
| Estimaciones de los Ods Ratios Y sus Intervalos de Confianza |            |                             |        |
|--|------------|-----------------------------|--------|
| Efecto   | Estimación | 95% Intervalos de Confianza |        |
| Clase 1 vs 3   | 14.861     | 7.635                       | 28.926 |
| Clase 2 vs 3   | 4.469      | 2.778                       | 7.191  |
| Femenino vs Masculino  | 11.818     | 8.006                       | 17.445 |
| Puerto de embarque C vs puerto de                            | 1.455      | 0.909                       | 2.328  |
| Puerto de embarque Q vs puerto de                            | 1.516      | 0.775                       | 2.966  |
| Edad   | 0.947      | 0.925                       | 0.97   |
| Cantidad de Familiares                                       | 0.758      | 0.588                       | 0.978  |
| Tarifa   | 1          | 0.996                       | 1.005  |

El odd ratio de supervivencia de la clase 1 fue 14.8 veces el odds ratio de supervivencia de la clase 3. Así mismo, el odd ratio de supervivencia de la clase 2 fue 4.469 veces el odds ratio de supervivencia de la clase 3.

En cuanto a las edades, a medida que aumenta la edad del pasajero aumenta el “likelihood” de no sobrevivir. Cada año Más que tiene el pasajero, este tiene un 94.7% del likelihood de sobrevivir que tenía un pasajero con un año menos de Edad. Esto mismo se puede decir para la tarifa pero en un 75.8%.

$$\text{Supervivencia} = \beta_1 \text{Clase} + \beta_2 \text{Sexo} + \beta_3 \text{Puerto de Embarque} + \beta_4 \text{Edad} + \beta_5 \text{Cantidad de Familiares} + \beta_6 \text{Tarifa}$$

Sacando las variables estadísticamente no significativas (**Tarifa y puerto de embarque**) la variable cantidad de familiares se hace significativa al 5%)

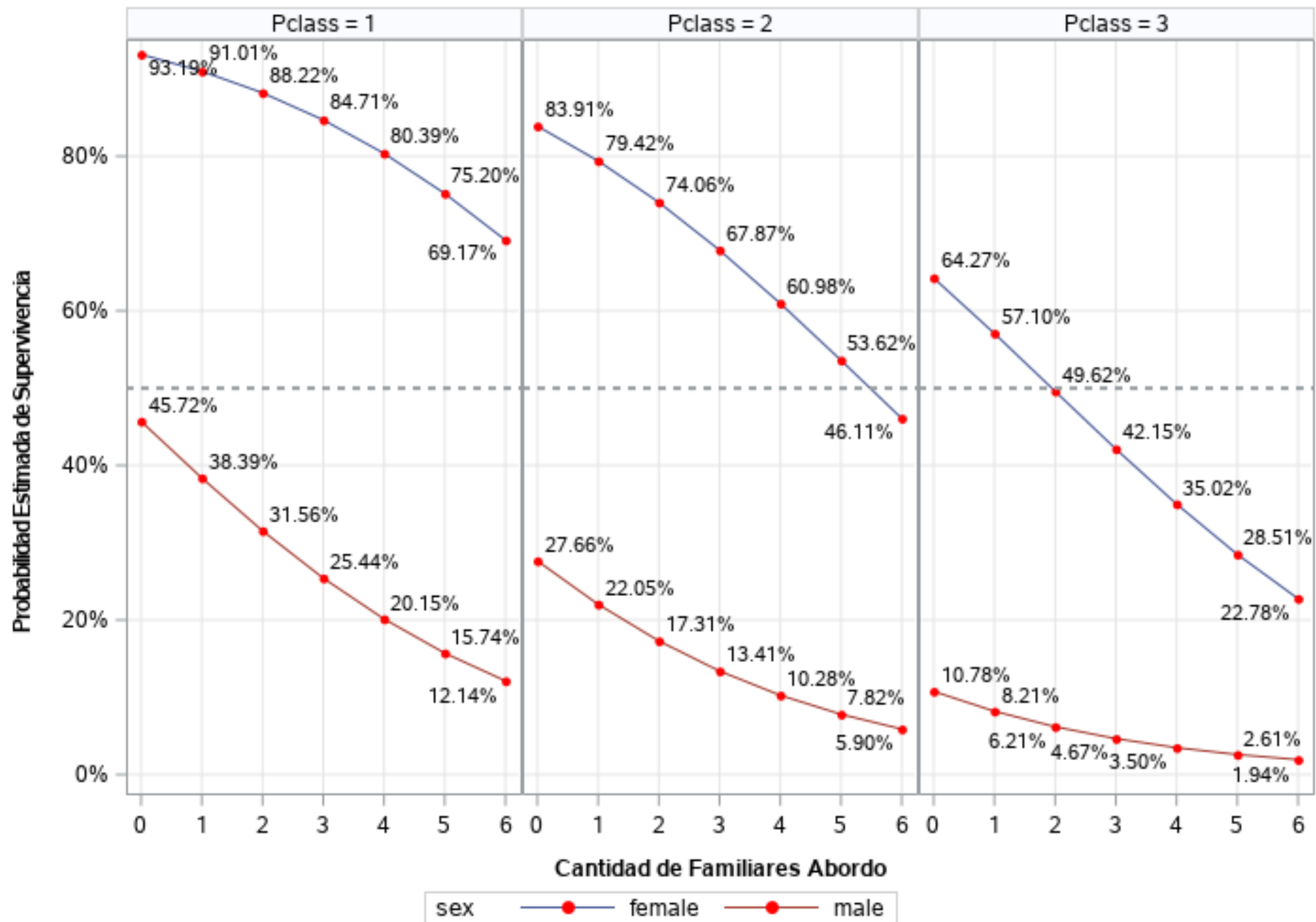


Manteniendo todo lo demás constante: numero de acompañantes=0



## Probabilidad Estimada de Supervivencia de un Pasajero

Dada la Cantidad de personas que viajaban con el/ella y la Clase



Manteniendo todo lo demás constante: edad mediana por sexo la clase en el grupo



# CONCLUSIONES

BASADAS EN EL MODELO



# EDAD Y SUPERVIVENCIA

- Los **más jóvenes** eran más likely a sobrevivir. Esto es cierto para todas las clases.





# SUPERVIVENCIA Y GÉNERO

- Las mujeres de todas las edades **eran mas likely a sobrevivir** que los hombres de todas las edades. Esto es cierto para todas las clases.





# SUPERVIVENCIA Y CANTIDAD DE FAMILIARES ABORDO

- Si se viaja con mas personas disminuye la probabilidad de supervivencia. Esto es cierto para todas las clases.





# TRATAMIENTO DE LOS VALORES VACÍOS

- Se imputo con el promedio la edad dada las características de genero, puerto de embarque, cantidad de familiares viajando.
- Ver diagnósticos del modelo adjunto.