

TRABAJO PRACTICO

Una compañía aérea desea determinar estudiar los factores que determinan la satisfacción/insatisfacción de sus clientes. Para eso realizó una encuesta donde relevó las características de los viajes e individuos para diferentes viajes y le solicitó a los pasajeros que indiquen si estaban o no satisfechos con sus viajes.

Las variables relevadas son las siguientes. Se utilizará como variable dependiente la variable *satisfecha* codificada como (1:si y 0:no).

satisfecho	Indica si el pasajero declaró estar satisfecho con la experiencia del viaje
sexo	Sexo del pasajero
lealtad	Lealtad del cliente de acuerdo con los registros de la línea aérea
edad	Edad del cliente de acuerdo con los registros de la línea aérea
tipo_viaje	Tipo de viaje: Personal o de Negocios
clase	Clase utilizada por el pasajero durante el vuelo
distancia	Distancia recorrida
min_demora_partida	Minutos de demora en la partida
minutos_demora_arribo	Minutos de demora en la llegada

1. Estime un modelo que incluya todas las variables disponibles.
 - a. Analice en términos del problema el signo de los coeficientes de las variables: sexo, tipo_cliente y clase. Realice este punto aún cuando el coeficiente de la variable no sea significativo (suponga que sí lo es).
 - b. Analice la significancia de las variables. Justifique el sentido de los test de hipótesis que plantea.
2. Utilice el procedimiento step hacia adelante y hacia atrás.
 - a. Analice la significancia de ambos modelos.
 - b. Evalúe los 2 modelos obtenidos. Compare los modelos mediante un único indicador para determinar cuál de los dos es mejor.
3. Estime un modelo con las siguientes variables explicativas: clase, distancia y minutos_demora_arribo.
 - a. Utilizando la expresión del modelo, grafique la probabilidad de que el pasajero se declare satisfecho en función de los minutos de demora del arribo (entre 0 y 800 minutos) para un viaje de una distancia igual a la distancia promedio. Realice una curva para cada una de las 3 clases.
 - b. Analice el gráfico anterior y extraiga conclusiones útiles para el negocio.
4. Divida la muestra que posee en una muestra de entrenamiento y otra de testing. Utilice las proporciones 75% / 25%. Utilice el comando `set.seed(123)` antes de hacer la partición.
 - a. Corra un modelo con las mismas variables utilizadas en el punto 3. Utilizando la muestra de Entrenamiento.
 - b. Calcule las probabilidades estimadas para todas las observaciones de los datos de Testing. Realice un histograma con dichos valores.
 - c. Utilizando los valores del punto 4b y una probabilidad de corte de 0,60, construya la matriz de confusión. Calcule la especificidad, la sensibilidad, la exactitud y la precisión. Interprete estos valores en términos del problema.