

### Caso de Aplicación 1: Sobreventa de asientos en Aerolíneas

Para esta pregunta considere el dataset caso\_noshow\_tarea.csv que se entrega junto con este enunciado. El dataset contiene información sobre sobre vuelos en la ruta Lima – Sao Paulo (en ambas direcciones), detallando:

- Fecha: Fecha del vuelo (año – mes – día)
- Vuelo: Numero de vuelo
- Dia\_sem: día de semana en que sale el vuelo (1= lunes, 7 = Domingo)
- Sem\_anio: semana del año correspondiente a la fecha de salida
- Asientos: capacidad de asientos del vuelo
- Pasajeros: Cuantos pasajeros se vendió ticket
- Hora\_despegue: hora planificada para el despegue
- Origen: aeropuerto de salida (LIM o SAO)
- Destino: aeropuerto de llegada (LIM o SAO)
- Pasajero\_conexion: cuantos pasajeros vienen de conexión
- Pasajeros\_negocio: cuantos pasajeros viajan en clase business
- Pasajeros\_noshow: cuantos pasajeros inscritos no llegaron a volar

Considere el siguiente vuelo:

- LATAM LIM-SAO
- Salida: 16 de octubre 2019, 1pm.
- Capacidad de avión = 150
- Precio ticket = US\$450, no reembolsable
- Costo de bajar pasajero = US\$1100
- Se vendieron todos los asientos del vuelo (150)

Responda las siguientes preguntas:

- a) Utilice los datos entregados para calcular la distribución de probabilidad empírica del número de no-shows para el vuelo especificado, asumiendo que el vuelo se vende por completo. Para esto, debe seleccionar una muestra de vuelos representativa, graficar un histograma que ilustre los no-show y luego usar la muestra para calcular la distribución de no show. Se recomienda analizar el porcentaje de no-shows de cada vuelo, usando el ratio  $(\text{pasajeros\_noshow})/\text{pasajeros}$ , y luego utilizar esto para calcular la distribución de no-show para el vuelo específico que se pide analizar.
- b) Considere ahora una distribución normal para analizar la distribución de probabilidad de los no-show. ¿Le parece adecuado usar esta distribución paramétrica? ¿Qué otras distribuciones se podrían utilizar para representar los no-show de forma paramétrica?
- c) Se entrega información adicional sobre el vuelo del 19 de octubre: (i) hay 10 pasajeros que viajan en business; y (ii) un 70% de los pasajeros tiene conexión. ¿Es útil esta información para mejorar la predicción no-shows del vuelo? De ser así: ¿cómo usaría esta información?
- d) Utilizando alguna de las distribuciones de no-show descritas en su análisis previo, encuentra la cantidad optima de asientos a sobrevender para el vuelo indicado.