

# Data-Driven Decision Making

Prof. Marcelo Olivares

Prof Ayudante Ian Malgarini

## Pregunta I – Sobreventa de asientos en Aerolíneas (25 pts)

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 (ano – 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

Además, se adjunta acá un link para ver los datos en una planilla Google.

<https://docs.google.com/spreadsheets/d/1CwStLcfMelzVF0G3SOyW7vRCbQ6PidY6rEgxbE71AUk/edit?usp=sharing>

### Considere el siguiente vuelo.

- LIM-SAO
- Salida: 16 de noviembre 2021, 1pm.
- Capacidad de avión = 150
- Precio ticket = US\$450, reembolsable
- Costo de bajar pasajero = US\$1100
- Se vendieron todos los asientos del vuelo (150)

El objetivo de esta tarea es usar los datos reales para calcular el numero de asientos a sobrevender en este vuelo de modo de maximizar el ingreso esperado.

### **Análisis exploratorio**

La primera actividad se centra en un análisis exploratorio de los datos. Notar que el número de tickets vendidos en cada vuelo en los datos históricos varía entre vuelos. Esto puede ser problemático para analizar el número de no-shows. Por ejemplo, no es lo mismo tener 5 no-shows en un vuelo que vendió 50 tickets vs. otro en donde se vendieron 130 tickets. Para abordar este problema, usamos la proporción de no shows en un vuelo:

$$NS\_Prop = \frac{\# NoShows}{\# Tickets\ vendidos}$$

Calcule y agregue esta columna en la planilla. Responda las siguientes preguntas:

- Determine si la proporción de no-shows varía entre horas del día. Comente sobre posibles explicaciones de los resultados obtenidos.
- Proponga otras dos variables que podrían afectar la proporción de no-shows. Utilice los datos para verificar o rechazar sus hipótesis.

### **Optimización de overbooking**

En el análisis exploratorio se analizó como la proporción de no-show varía entre vuelos de acuerdo a ciertas características. Tomando esto en consideración, responda las siguientes preguntas:

- Seleccione un subconjunto de vuelos que sea adecuado para representar el comportamiento de los no-show del vuelo que nos interesa definir el overbooking. Usando esta muestra, genere un histograma que represente la incertidumbre en el número de no-shows para el vuelo de interés.
- Basado en el histograma construido, determine el número de asientos a sobrevender en el vuelo.

## Pregunta II – Estimación de elasticidad de demanda (25 pts)

En esta pregunta utilice el dataset *e\_Car\_Data\_Extract\_forClass.xlsx*. Se adjunta además un tutorial y los códigos R que ilustran los análisis vistos en clases. (Puede realizar su análisis usando cualquier software: Python, R, Excel u otro).

- a) Con los datos proporcionados, estime un modelo lineal y un modelo logit que permita estimar la propensión a aceptar un crédito dado un APR. Explique las variables incluidas en el modelo y justifique. Despliegue los resultados en una tabla.
- b) Utilizando alguno de los modelos estimados, calcule la tasa optima (APR) a ofrecer a un cliente con FICO 700, para un monto de \$22000 y un costo de fondo del crédito de CF=5.8. Para realizar el este análisis, considere que la utilidad que genera el crédito esta dada por:

$$Utilidad = Monto \times (R - CF) \times G(R)$$

En donde:

Monto = monto del crédito

R = tasa ofrecida (= APR, la variable de decisión)

CF = Costo de fondo

G(R)= Probabilidad que el cliente acepte el crédito si la tasa ofrecida es R.

Si incorpora otras variables en su modelo estimado, puede fijar valores “razonables” para realizar el análisis.

### Pregunta III – Revenue Management (20 pts)

- 1) En la planilla Excel *data\_crucero.xlsx* se muestran las valoraciones (máxima disposición a pagar) de los clientes que han respondido la encuesta de cuanto valoran el crucero.

Basado en estos datos:

- Construya la curva de demanda, que indique la probabilidad que un cliente compre un paquete que se vende a un precio  $P$ . Para esto, debe calcular la fracción de clientes que tiene una valoración igual o menor a  $P$ , para distintos valores de  $P$ .
  - Usando la curva de demanda calculada, calcule el precio óptimo que maximiza el ingreso esperado cuando tiene un número ilimitado de paquetes para vender.
  - Considere ahora el caso en que vende durante 5 días seguidos, y cada día llegan en promedio 15 clientes. Se busca maximizar el ingreso esperado. ¿Que precio fijaría el primer día si solo cuenta con 30 paquetes para vender durante los 5 días?
  - Esta por iniciarse el quinto día de ventas y le quedan 10 piezas. ¿Que precio fijaría si su objetivo es maximizar el ingreso esperado?
- 2) Considere ahora la última versión del juego que hicimos en clase, en donde cuenta con 40 piezas para vender durante 5 días. Los primeros 4 días llegan clientes de baja valoración y el último día clientes con alta valoración. Considere los siguientes parámetros:
- El precio en los primeros 4 días se fija en \$ 500.000 y el último día a una tarifa alta de \$ 1.300.000.
  - El número de clientes que compra el producto *el último día* (cuando se ofrece la tarifa alta) sigue una distribución de Poisson con media 12 (la tabla mas abajo muestra esta distribución de demanda).

¿Cuándo paquetes se debería proteger para el último día si el objetivo es maximizar el ingreso esperado de los 5 días?

Hint: para este problema no se necesita utilizar la curva de demanda que se calculó en la pregunta anterior.

Tabla Distribucion de Poisson con media =12

x	Pr(D=x)	Pr(D≤x)	x	Pr(D=x)	Pr(D≤x)
1	0.00	0.00	11	0.11	0.46
2	0.00	0.00	12	0.11	0.58
3	0.00	0.00	13	0.11	0.68
4	0.01	0.01	14	0.09	0.77
5	0.01	0.02	15	0.07	0.84
6	0.03	0.05	16	0.05	0.90
7	0.04	0.09	17	0.04	0.94
8	0.07	0.16	18	0.03	0.96
9	0.09	0.24	19	0.02	0.98
10	0.10	0.35	20	0.01	0.99

### Problema IV: Tiempos de Espera (15 pts.)

FestEvents es un organizador de conciertos que está evaluando la instalación de baños portátiles en un lugar al aire libre donde se llevará a cabo un concierto. En particular, está considerando un área del lugar que está lejos de los baños existentes. Supongamos que habrá una sola fila para acceder a todos los baños portátiles en esta área y que solo una persona usa un baño a la vez. FestEvents está preocupado de que si la fila es demasiado larga, las personas encontrarán otras formas y lugares para responder a la llamada de la naturaleza. FestEvents estima que habrá un promedio de 4 llegadas por minuto a los baños, y que estas llegadas siguen un proceso de Poisson. El tiempo que una persona pasa usando el baño sigue una distribución exponencial con media 3.12 minutos.

1. Supongamos que FestEvents instala 14 baños portátiles. En promedio, ¿qué fracción de los baños estará ocupada durante el evento?
2. Si se instalan 14 baños: ¿Cuál es el tiempo promedio de espera?
3. ¿Cuál es el número mínimo de baños portátiles que FestEvents debe instalar para asegurarse de que la longitud de la fila no siga creciendo sin límite (es decir, para que la fila sea estable)? La respuesta debe ser un número entero.

## Problema V: Newsvendor (15 pts)

Montanso es una empresa biotecnológica que produce semillas para la agricultura. Necesitan decidir cuántas semillas producir para satisfacer la demanda de la próxima temporada. A Montanso le cuesta \$8 producir cada kilogramo (kg) de semilla, con un precio de venta de \$45 / kg. Si produce más semillas de la que demandan los agricultores locales, las semillas restante se envía al extranjero. El precio de venta en el mercado internacional es menor, \$3 por kg. (pero esto es mejor que destruir la semilla porque no se puede almacenar hasta el próximo año). Si la demanda supera la cantidad producida, entonces se pierden las ventas (los agricultores acuden a otro proveedor). Como pronóstico de la demanda, utilizarán una distribución normal con una media de 300.000 y una desviación estándar de 106.000.

1. ¿Cuántos kilogramos deberían producir para maximizar las utilidades esperadas durante la temporada?
2. ¿cuántos kilogramos deberían producirse si desean minimizar su inventario asegurando que la probabilidad de quedarse sin stock no sea mayor al 10%? (esta pregunta no requiere responder la pregunta 1 para ser analizada)
3. Un alto ejecutivo de la empresa hace la siguiente pregunta: "Supongamos que producimos 500,000 kilogramos ¿Cuál es la probabilidad de que nuestros ingresos totales sean al menos \$18,000,000?" Tomar en cuenta que los ingresos provienen tanto de las ventas locales como de las ventas en el extranjero. Además, se pregunta por los ingresos y no las utilidades, por lo que se debe ignorar los costos.