

Ciencia de Datos en Multinacional de Supermercados



Supongamos que trabajas para una empresa multinacional como Walmart. Se desea abrir una nueva sucursal en una comunidad como Juriquilla con un total aproximado de 40,000 habitantes.

El documento .csv se te proporcionó para hacer tus estimaciones. Es un documento que cuenta con las ganancias por transacciones en comunidades semejantes cuanto a nivel socioeconómico. Por ganancia se refieren a que esta cantidad de dinero ya contempla

$$\text{Ganancia} = \text{Transacción} - \text{Pago de Impuestos} - \text{Costos de Compra de Productos}$$

Sin embargo, queremos saber el porcentaje de personas que necesitamos convencer a que compren en nuestra nueva sucursal para tener una ganancia neta mensual de al menos 1,500,000.

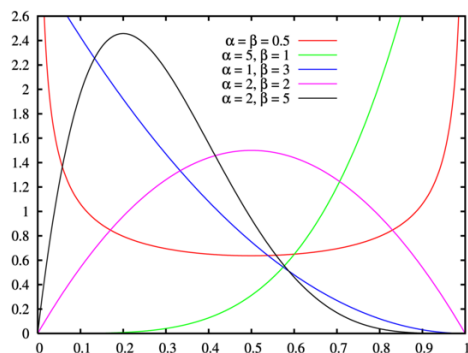
Para poder lograr nuestro objetivo necesitamos hacer estimación de parámetros. Uno de los métodos más populares es el llamado en inglés **máximum likelihood estimation**.

A la par utilizaremos una de las distribuciones más generales y flexibles que existen, la cual es la función de densidad de probabilidad beta. Esta función de densidad tiene la gran ventaja de que al variar sus parámetros α y β , se puede ajustar fácilmente a diferentes formas de distribución de los datos.

$$f_X(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$$

Donde la función de densidad de probabilidad está definida de 0 a 1 , y fuera de este rango es 0 y

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx$$



Por lo que es razonable asumir que la distribución de los datos sigue una función de densidad beta. Lo que necesitamos es determinar α y β .

Los pasos a seguir son los siguientes

1. Estudia que es el algoritmo de maximum likelihood
2. Una vez entendido un estimador por maximum likelihood, investiga como fácilmente podemos estimar los parametros α y β de una función densidad beta con los datos que nos dan en Python. Una vez logrado esto, la función de densidad encontrada modela (al integrarla) la probabilidad de ganancia en un cierto intervalo. NOTA: La función densidad solo acepta datos entre 0 y 1. Por lo que necesitamos normalizar los datos. Los normalizaremos como

$$X_{norm} = \frac{1}{max - min} (X - min)$$

donde max es la valor máximo de los datos y min es el valor mínimo de los datos.

3. Una vez que tienes los parametros α y β estimados, necesitamos calcular la desviación estándar σ y la media μ . Investiga como determinar estos dos valores para una distribución beta.
4. Ahora deseamos saber si $S_n = X_1 + X_2 + X_3 + \dots + X_n$ (dinero ganado por N ventas) logrará ser al menos mayor de \$1,500,000+ gastos de operación. Para determinar los gastos de operación de la sucursal te mencionan lo siguiente:
Para los almacenistas, las personas que atienden en el supermercado, los cajeros y conserjes usa las siguientes fuentes

https://www.glassdoor.com.mx/Sueldos/conserje-sueldo-SRCH_KO0,8.htm

https://www.gob.mx/cms/uploads/attachment/file/873886/Tabla_de_Salarios_M_nimos_2024.pdf . Deseamos pagarles 15% más del salario mínimo de cada caso.

Para los almacenistas, las personas que atienden en el supermercado, los cajeros considera que cada uno trabaja 5 veces a la semana.

- Necesitamos 40 personas en almacén
- Necesitamos 60 personas que atiendan a clientes en los pasillos del supermercado. (para fines de pago los consideraremos como personas de mostrador)
- Necesitamos 30 cajeros
- Necesitamos 20 conserjes.
- Necesitamos 1 gerente general a quien se le pagarán 100,000.00 pesos mensualmente.
- Necesitamos 4 subgerentes a quienes se les pagarán en 45,000.00 pesos mensualmente a cada uno.
- Nuestra sucursal tendría una superficie de 2000 m^2 , se gasta aproximadamente $120 \frac{\text{kW}}{\text{hr}} \frac{1}{\text{m}^2}$ en nuestras sucursales. La sucursal estaría abierta de 10:00 a.m. a 10:00 p.m. La CFE cobra aproximadamente \$2.3 por cada kilowatt*hora.

Con esta información calcula el número de ventas que necesitamos hacer para que en un mes logremos cubrir los gastos de operación más una ganancia de \$1,500,000 pesos mensuales con un 99% de probabilidad. NOTA: TEN CUIDADO. Lo que nos interesa es hacer estimaciones con las variables aleatorias X NO X_{norm} . Haz todo el cálculo en Python.

5. Una vez estimada la cantidad de ventas, supongamos que una compra (para nosotros como negocio es ventas) es realizada una vez por habitante en la comunidad semejante a Juriquilla por semana. Estima el porcentaje de los habitantes de esta comunidad que debemos convencer. Haz el cálculo en Python.

ANALIZANDO LOS RATINGS

6. Ve la distribución de los ratings
7. ¿Qué distribución sigue aproximadamente?
8. ¿Cuál es la probabilidad de que en la sucursal que abramos, si los datos que usamos fue en una comunidad semejante a la de Juriquilla, de que en promedio tengan un rating de 8.5 o más? ¿Qué le dirías a los del departamento de atención a clientes? (USA TEOREMA DEL LÍMITE CENTRAL DE NUEVO)

MEJORA EL ANALISIS

PUNTOS POSIBLES A MEJORAR

- Luz (el gasto que usamos fue encontrado en Google). Busca alguna referencia de CFE o alguna referencia más confiable.
- ¿Faltaron gastos por contemplar? ¿Agua? ¿Luz 24 horas?
- Número de Empleados. ¿Nos faltaron trabajos por contemplar?
- Asumimos que una persona va al supermercado una vez por semana... ¿esto es cierto? ¿Qué otra cosa podrías hacer/asumir?

