

UNIVERSIDAD ADOLFO IBÁÑEZ

FACULTAD DE INGENIERÍA Y CIENCIAS

MAGISTER Y DIPLOMADO EN CIENCIA DE DATOS

TALLER PRÁCTICO – Estadística para Data Science

Profesor: Alonso Molina

Ayudante: Felipe Salazar

Se posee la información de cierta cadena de construcción y artículos para el hogar, con presencia en 17 SECTORES del país y con 109 sucursales de una tienda que clasifica a sus productos en 4 categorías (TIPO_ARTICULO). El número de cuenta (ACCOUNTID) está asociado a cada cliente (que podría comprar al menos una vez). El MONTO_PAGADO corresponde al total de la compra del TIPO_ARTICULO según la CANTIDAD_ARTICULO comprada, mientras que el MARGEN es por artículo.

Cada fila representa un solo artículo, del cual puede llevar varias unidades y por tanto, el valor unitario se puede obtener directamente. Por ejemplo, si una fila dice que la cantidad es 2 artículos, entonces asuma que los 2 artículos valen lo mismo.

Por otro lado, el DESCUENTO es para el total de la compra, no por artículo.

Usted trabajará solo con el SECTOR asociado a su GRUPO.

1. [5 puntos] Obtenga una tabla que entregue la probabilidad de utilizar cierto METODO_DE_PAGO, dado el TIPO_ARTICULO, y analice los resultados.
2. [5 puntos] Obtenga el valor unitario de los productos y realice un análisis descriptivo univariado de todas las columnas. Puede utilizar "EDAs automáticos" para simplificar la programación.
3. [5 puntos] Realice un tratamiento de datos: determine datos atípicos, NA y tome acciones con ellos.
4. [5 puntos] Analice la correlación entre el Margen y el Valor unitario. Considere el tipo de artículo en su análisis, y realice hipótesis al respecto, utilizando un 5% de significancia.
5. [5 puntos] Un experto cree que el número de compras por día sigue una distribución Poisson. Obtenga un resumen de compras por día, **estime el parámetro** y compruebe la probabilidad empírica y teórica de que un día cualquiera se realicen a lo más 30 compras.
6. [5 puntos] Proponga una distribución para el valor unitario de los productos de Calidad Alta y, con un 5% de significancia, determine si es correcta su propuesta. Utilice un gráfico para acompañar su propuesta.
7. [5 puntos] Obtenga un intervalo de confianza para la **media** de descuentos utilizados, y luego un intervalo de confianza para la **proporción** de compras donde se aplicó un descuento. Elija la confianza.
8. [5 puntos] Compare las medias y varianzas del margen entre la Calidad Media y Alta y concluya, con un 5% de significancia, si son iguales o distintas.

Instrucciones

Deberá entregar en formato PDF o HTML su resultado, sin códigos a la vista,

El trabajo se realiza con el grupo inscrito en el foro disponible de webcursos.

Debe enviar la respuesta en formato **.R** al correo alonso.molina@edu.uai.cl y a felipe.salazar@edu.uai.cl a más tardar el día viernes 18 de julio a las 23.59 hrs.