



FACULTAD DE MATEMÁTICAS
PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

DEPARTAMENTO DE ESTADÍSTICA
EYP3007 - Consultoría Estadística

Reporte: Tarjeta de Compra Más (TCM)

*Eduardo Vásquez
Emiliano Moreno
Ricardo Torres*

Profesor: Ricardo Aravena

Santiago de Chile, Abril del 2022

1 Introducción

El problema a resolver en el siguiente informe es poder dilucidar y entender los problemas que presenta la Tarjeta de Compra Más. En una reunión entre las diferentes partes de la empresa, se presentaron cuatro hipótesis sobre posibles problemas en la TCM.

A continuación se presenta una pequeña vista a los datos entregados, para luego explicar la metodología a utilizar para estudiar las hipótesis planteadas. Luego, en la sección **Hipótesis a Cubrir**, se presentan y se estudian estas hipótesis, siguiendo la metodología descrita.

Finalmente, en la sección **Conclusión**, se entrega un resumen de los resultados obtenidos en la sección anterior, y al mismo tiempo se proponen diferentes medidas para mejorar el uso de la TCM.

1.1 Datos entregados

Para los estudios solicitados, el Departamento de Informática nos hizo llegar una pequeña muestra aleatoria de 450 clientes, la cual contiene información histórica y los atributos necesarios para nuestro trabajo. Entre estas variables se encuentra el ID del cliente asignado por la empresa, la edad, el sexo y la región del cliente, así como métricas de uso de la TCM, en particular si se utilizó la tarjeta el primer trimestre y el monto de compras durante este mismo periodo.

1.2 Metodología

La metodología a utilizar consiste en:

1. Presentar y formalizar cada una de las hipótesis.
2. Realizar un análisis gráfico de los datos pertinentes al problema.
3. Aplicar un test estadístico adecuado.

2 Hipótesis a Cubrir

2.1 Bajo uso

La primera hipótesis planteada es que el porcentaje de uso de la TCM, en el primer trimestre del 2022, ha disminuido con respecto al uso en el 2021, que fue de un 62%.

A partir de los datos vemos que el porcentaje de uso en el año 2022 corresponde a 58%. Podemos ver una comparación entre ambas cantidades en la Figura 1.

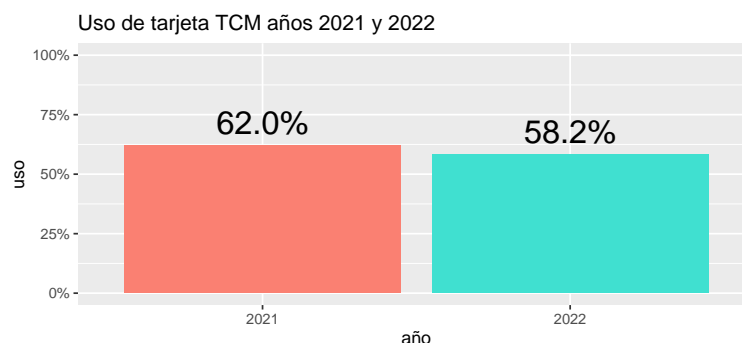


Figura 1: Porcentaje uso tarjeta años 2021 y 2022

Se puede observar en el gráfico que efectivamente el uso de la tarjeta parece ser menor para el primer trimestre del 2022, en comparación con el mismo periodo del año 2021.

Ahora, proponemos un test binomial exacto para confirmar si los hallazgos son significativos. Este test es adecuado pues permite establecer si una muestra de datos binarios proviene de una población con proporción menor a un valor dado.

Específicamente, tenemos que nuestra hipótesis nula es que el porcentaje de uso *no* ha disminuido. El test nos entrega un valor-p de 0.055, así, considerando un nivel de significancia de $\alpha = 0.05$, no tenemos evidencia suficiente para rechazar esta hipótesis. Presentamos algunos comentarios adicionales en la conclusión del informe.

2.2 Montos

En esta sección se estudia la hipótesis de que el monto promedio utilizado en el mes de Marzo de 2022 es mayor al monto promedio utilizado en el mismo mes del año 2021.

Para el cálculo del monto promedio utilizado en marzo de 2022, solo se considera el monto tranzado por clientes que sí registran uso en su tarjeta para este mes. Se obtiene así que el monto promedio en marzo del 2022 es 449.3. Ahora se verificará que la diferencia encontrada es estadísticamente significativa.

Una alternativa es utilizar un *test-t*, sin embargo para que los resultados sean válidos se debe asegurar que los datos provienen de una distribución Normal.

En la figura 2 se estudia el supuesto de normalidad de manera gráfica mediante estimación de densidad y gráfico QQ-Norm.

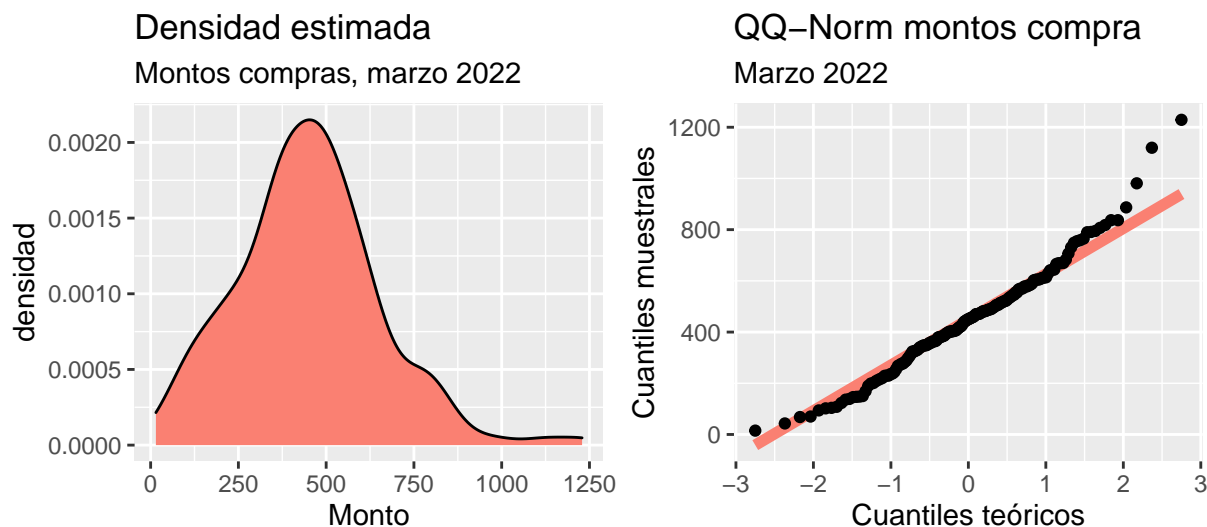


Figura 2: Test gráficos para normalidad media montos compra

Ahora se complementa el análisis gráfico con un test específico para normalidad: Shapiro-Wilk

Para el test de Shapiro-Wilk se tiene un Valor-P de 0.0087. En consecuencia, no podemos asumir normalidad, por lo que ocupamos un test no paramétrico (Wilcoxon) para establecer la significancia en la diferencia entre el promedio de los montos tranzados en Marzo de 2022 y 2021.

Para el test de Wilcoxon se tiene un Valor-P de 0.0021. En consecuencia, aceptamos la hipótesis que el promedio de compras en marzo de 2022 fue mayor que en marzo de 2021.

2.3 Antiguos

En esta sección se estudia la hipótesis de que, entre quienes utilizaron su tarjeta, los montos medios tranzados por clientes antiguos es cien mil pesos menor que el de clientes nuevos. Para eso, segmentamos en clientes antiguos y nuevos. Tampoco consideramos clientes que no registran uso de su tarjeta en el periodo examinado.

Tenemos que hay 67 clientes antiguos y 195 clientes nuevos presentes en la base de datos entregada. Además, los montos medios de cada grupo son 1114.2985 y 1242.0154, respectivamente.

Queremos ocupar un test t para dos muestras. En la figura 3 estudiamos si los datos siguen una distribución Normal.

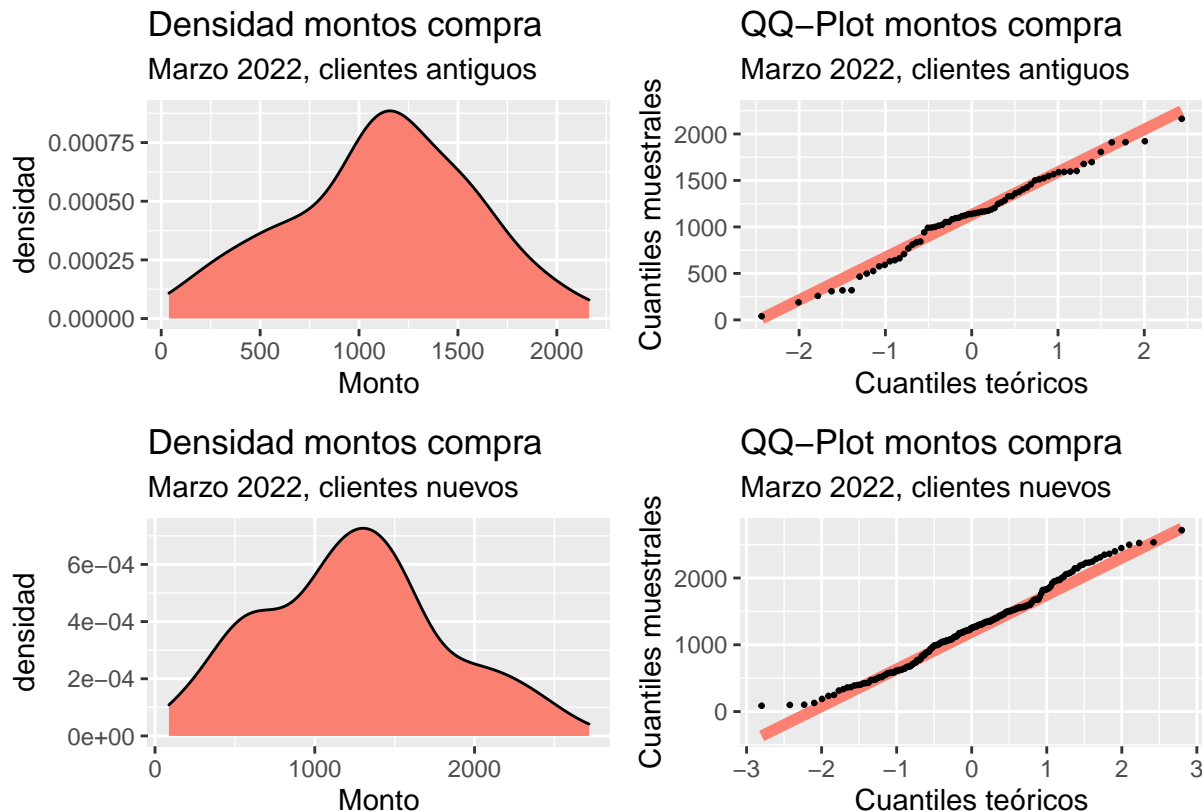


Figura 3: Test gráficos de normalidad media montos compra por segmento

Como en la sección 2.2, se confirman los hallazgos con el test de Shapiro-Wilk. Específicamente, se obtuvieron valores-p iguales a 0.5825 y 0.0491 para clientes antiguos y nuevos respectivamente. En los clientes nuevos además utilizamos el test de Kolmogorov-Smirnov al obtener un valor-p = 0.6882.

Como no hay evidencia para rechazar normalidad, se aplica un *test-t* para estudiar la diferencia de medias, asumiendo varianzas desconocidas. Tenemos que ver si usamos varianzas iguales y desconocidas o diferentes y desconocidas mediante la función de R `var.test`. Se rechaza dicho test con un valor-P = 0.0463, por lo que trabajamos varianzas diferentes.

Para el *test-t* se tiene un valor-P = 0.3465. Luego, no es posible rechazar la hipótesis de que el promedio de gasto de clientes antiguos es cien mil pesos menor que el de clientes nuevos.

2.4 Atributos

En esta sección se estudia la hipótesis de que hay un comportamiento diferente en frecuencia de uso de la tarjeta y los montos transados, en función de la edad (como categoría), el sexo o región. Para esto, primero graficamos para cada categoría la distribución de los montos acumulados (figura 4). Por último se realiza un análisis *anova* para cada categoría, utilizando como variable respuesta el monto acumulado y el uso de la tarjeta.

Según el análisis, hay evidencia para rechazar la hipótesis solamente para la frecuencia del uso de la tarjeta

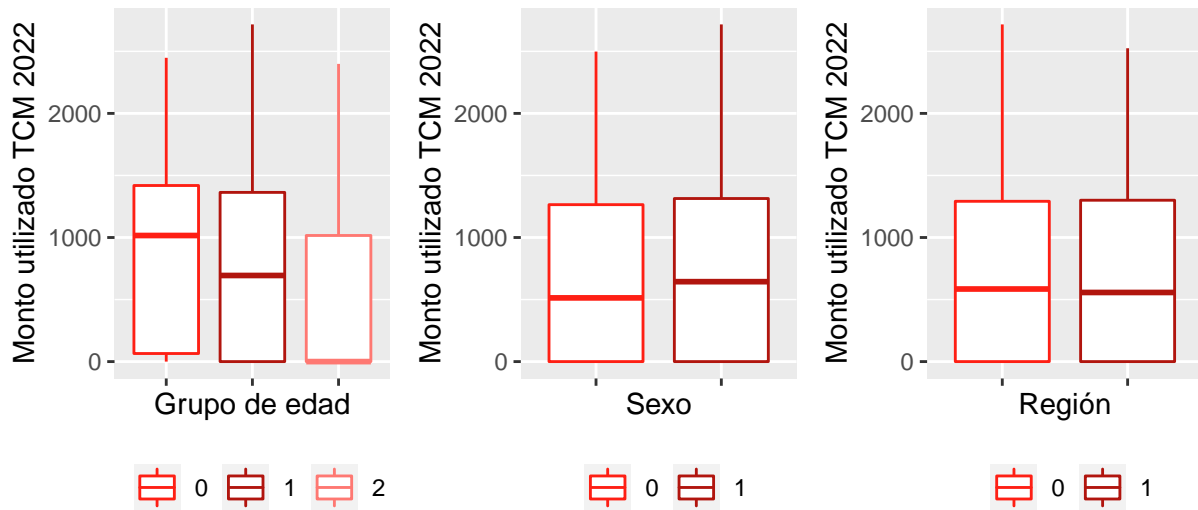


Figura 4: Box-Plot categorías vs monto acumulado

en función de la edad.

3 Conclusión

Obtuvimos que

1. No tenemos evidencia suficiente para concluir que el porcentaje de uso de la tarjeta ha disminuido. De todas maneras, creemos que en este caso podemos relajar el nivel de significancia y declarar que hay evidencia.
2. Sí tenemos evidencia estadística para declarar que el comité de promociones ha focalizado apropiadamente las ofertas.
3. En cuanto a la acusación al área de fidelización, obtenemos que no existe evidencia estadística suficiente.
4. En cuanto a los atributos, tenemos evidencia suficiente para decir que la frecuencia de uso de la tarjeta depende de la edad de la persona, por lo que deberíamos focalizar las promociones en los grupos que menos las utilizan

Así, recomendamos a la empresa que ...