



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 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, y se explica la metodología a utilizar para estudiar las hipótesis planteadas. Luego, en la sección **Hipótesis a Cubrir**, se presentan y se estudian las cuatro hipótesis planteadas durante la reunión de la empresa.

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

La base de datos con la que se trabaja corresponde al registro de 450 clientes para los cuales se tienen los siguientes atributos:

- Cliente: Identificador único
- Edad:
- Sex:
- Reg:
- Uso2022:
- UsoMarzo
- MontoMarzo
- MontoAcum

1.2 Metodología

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 corresponde a 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.



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

Ahora, proponemos un test binomial exacto para confirmar si 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.

Si relajamos algunas exigencias, en particular el nivel de significancia a un 10%, concluimos que hay evidencia en favor de que la disminución en el uso de la tarjeta es estadísticamente significativa.

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 esto, primero calculamos este promedio para cada mes. Notamos que existe una cantidad importante de clientes con monto 0. Esto se debe a que hay clientes que no utilizan su tarjeta durante Marzo 2022. Para este análisis, solo se considera el monto tranzado por clientes que sí registran uso en su tarjeta para este mes.

```
monto_promedio_marzo_2021 <- 400

monto_promedio_marzo_2022 <- datos_tcm_compras %>%
  dplyr::summarise(media = mean(MontoMarzo)) %>%
  as.numeric()
```

Obtenemos 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 las siguientes figuras se estudia el supuesto de normalidad de manera gráfica mediante estimación de kernel y gráfico QQ.

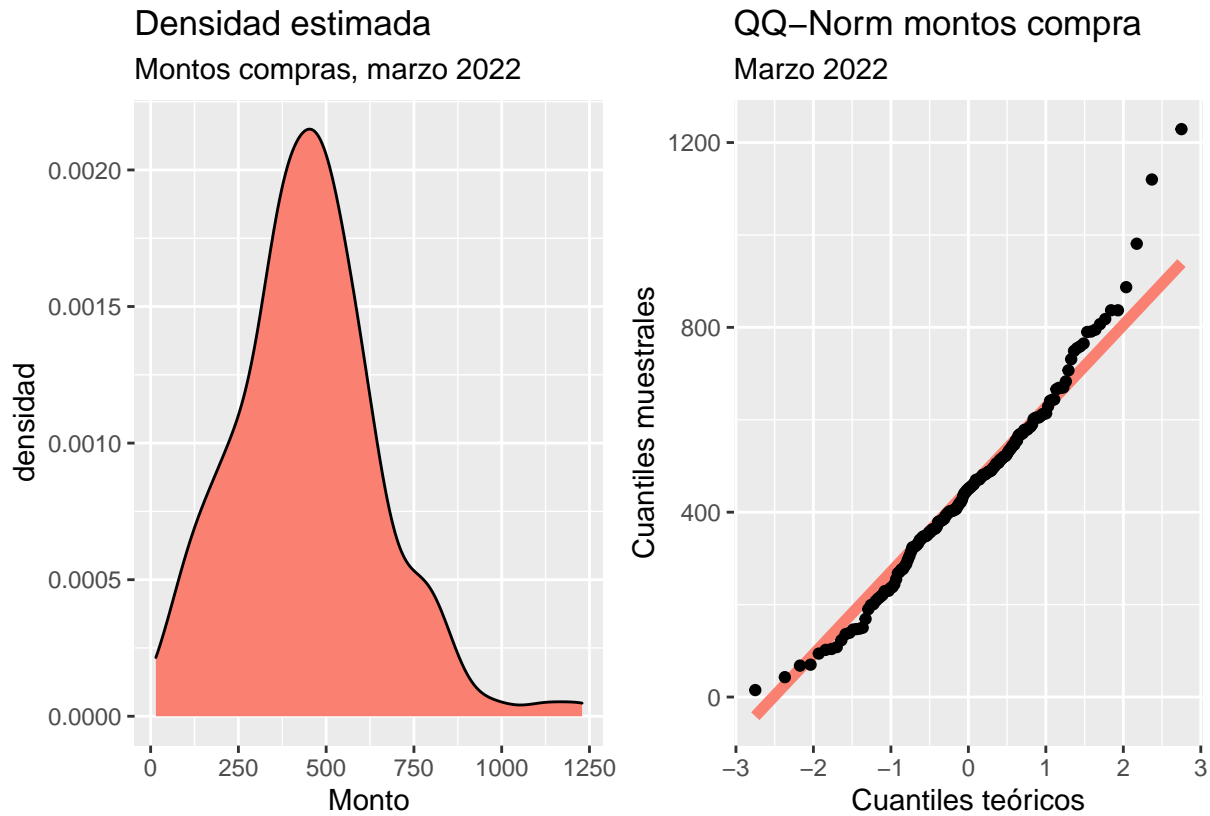


Figura 1: Test gráficos para normalidad media montos compra

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, encontramos que hay evidencia en favor de que la diferencia es significativa.

Ahora, en favor de la comparación, aplicamos un test t. Si asumimos que el tamaño de muestra es lo suficientemente grande, la aplicación de este test se valida con resultados asintóticos. Se observan resultados consecuentes en ambos test, pues para el *test-t* el valor -P es 0.001

Vemos que en ambos casos rechazamos la hipótesis que el monto del 2022 es menor al del 2021, por lo que aceptamos que fue superior al del 2021...

2.3 Antiguos

En esta sección se estudia la hipótesis de que la diferencia en los montos medios tranzados por clientes antiguos y nuevos es de cien mil. Para eso, segmentamos en clientes antiguos y nuevos. De manera similar como en la sección 2.2, no 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, así que vemos normalidad...

Ahora se estudia si los datos siguen una distribución Normal mediante métodos gráficos: primero por estimación kernel y luego mediante QQ-plot.

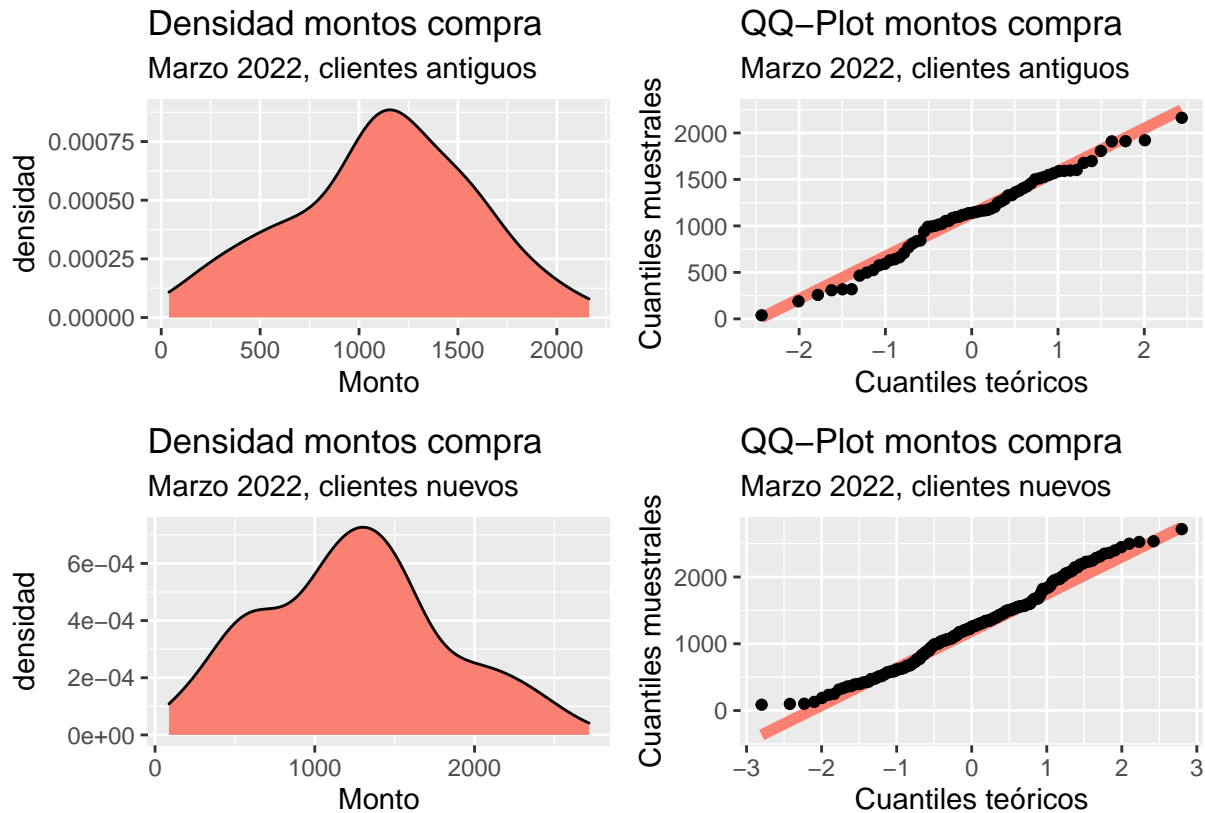


Figura 2: Test gráficos de normalidad media montos compra por segmento

De manera similar como en la sección 2.2, se confirman los hallazgos con test formales: usamos Shapiro-Wilk. En los clientes nuevos además utilizamos el test de Kolmogorov-Smirnov al obtener un valor-p bastante cercano a 0.05...

No se encuentra evidencia para rechazar normalidad. En consecuencia se aplica un *test-t* para estudiar la diferencia de medias.

Primero tenemos que ver si usamos varianzas iguales y desconocidas o diferentes y desconocidas.

Se rechaza el test con un valor-P 0.0463, por lo que trabajamos varianzas diferentes. Se asume que las varianzas son desconocidas.

Vemos entonces que no podemos rechazar la hipótesis nula...

2.4 Atributos

Agrupamos los datos de edades en los grupos correspondientes...

Acá rellenamos con gráficos si no nos pasamos de las 5 páginas...

Acá van los ANOVAS...

Concluimos

3 Conclusión