

Desarrollo Trabajo Práctico #3

Leandro Lepera

Marck Murillo

Yandri Uchuari

December 8, 2024

1 Planteamiento de problemas

El pequeño negocio de Don Francisco ha crecido considerablemente, pero hay incertidumbres sobre el rendimiento comparativo de las cinco sucursales actuales ('Santa Ana', 'La Floresta', 'Los Cedros', 'Palermo' y 'Córdoba'). Las inquietudes que se buscan abordar son:

1. ¿Cuál es la estabilidad de las ventas del supermercado 'Santa Ana' durante los últimos meses? ¿Qué tan confiable es esta información?
2. ¿Las ventas promedio entre todas las sucursales son homogéneas, o hay diferencias significativas que se deben atender?
3. ¿Qué tan diferente es el rendimiento de la tienda con mayor promedio de ventas frente a la que tiene el menor promedio? ¿Esta diferencia es estadísticamente significativa?

2 Procesamiento de los datos

Para los datos la manipulacion de los datos se siguieron los siguientes pasos:

1. se uso la funcion ExcelFile para cargar la hoja de calculo.
2. se inicializo la variable sheet_names con los nombres de las hojas.
3. se empleo la funcion de la libreria pandas read_excel usando el nombre de cada hojas para acceder a los arreglos de datos.

3 Resultados

3.1 problema 1. Intervalos de Confianza por mes

Análisis de ventas - Tienda Santa Ana

Mes	Ventas promedio	IC 95%	IC 99%
Enero	\$16,311.30	[\$15,223.36, \$17,399.25]	[\$14,846.34, \$17,776.26]
Febrero	\$17,901.85	[\$16,716.43, \$19,087.26]	[\$16,301.13, \$19,502.57]
Marzo	\$20,169.85	[\$19,189.30, \$21,150.41]	[\$18,849.50, \$21,490.21]
Abril	\$18,433.37	[\$17,396.55, \$19,470.18]	[\$17,036.04, \$19,830.70]
Mayo	\$20,164.56	[\$19,202.20, \$21,126.93]	[\$18,868.70, \$21,460.43]
Junio	\$21,093.06	[\$19,964.05, \$22,222.07]	[\$19,571.47, \$22,614.64]
Julio	\$20,208.69	[\$19,195.82, \$21,221.55]	[\$18,844.82, \$21,572.55]

Agosto	\$21,242.42	[\$20,217.34, \$22,267.50]	[\$19,862.11, \$22,622.73]
Septiembre	\$21,538.02	[\$20,545.93, \$22,530.11]	[\$20,200.96, \$22,875.07]
Octubre	\$21,185.50	[\$20,193.41, \$22,177.58]	[\$19,849.61, \$22,521.38]
Noviembre	\$21,177.24	[\$20,019.88, \$22,334.60]	[\$19,617.45, \$22,737.04]
Diciembre	\$18,753.83	[\$17,832.00, \$19,675.65]	[\$17,512.55, \$19,995.10]

Resultados por Mes

Mes	Media	Amplitud 95%	Amplitud 99%
Enero	\$16,311.30	\$2,175.89	\$2,929.92
Febrero	\$17,901.85	\$2,370.83	\$3,201.44
Marzo	\$20,169.85	\$1,961.12	\$2,640.72
Abril	\$18,433.37	\$2,073.63	\$2,794.66
Mayo	\$20,164.56	\$1,924.74	\$2,591.73
Junio	\$21,093.06	\$2,258.02	\$3,043.17
Julio	\$20,208.69	\$2,025.73	\$2,727.73
Agosto	\$21,242.42	\$2,050.16	\$2,760.61
Septiembre	\$21,538.02	\$1,984.18	\$2,674.11
Octubre	\$21,185.50	\$1,984.17	\$2,671.76
Noviembre	\$21,177.24	\$2,314.72	\$3,119.58
Diciembre	\$18,753.83	\$1,843.66	\$2,482.55

Análisis de Variabilidad

Para Intervalos del 95%

- **Menor variabilidad:** Diciembre (Amplitud: \$1,843.66)
- **Mayor variabilidad:** Febrero (Amplitud: \$2,370.83)

Estadísticas Generales

Estadística	Valor
Media anual de ventas	\$19,856.50
Desviación estándar	\$3,181.05
Ventas mínimas	\$11,222.63
Ventas máximas	\$28,793.36

3.2 problema 2. Prueba ANOVA para Comparación de Ventas Promedio entre Tiendas

Planteamiento de la Prueba de Hipótesis

Se define un nivel de significancia del $\alpha = 0.05\%$ y se plantea la siguiente prueba de hipótesis:

H_0 : Las ventas promedio de todas las tiendas son estadísticamente iguales.

H_1 : Existe al menos una tienda cuyas ventas promedio son estadísticamente diferentes.

Método y Resultados

La prueba se realizó utilizando la función `f_oneway` de la librería `scipy`. A continuación, se presentan los resultados obtenidos:

- Estadístico F : 74.2501
- Valor p : 2.3354×10^{-58}

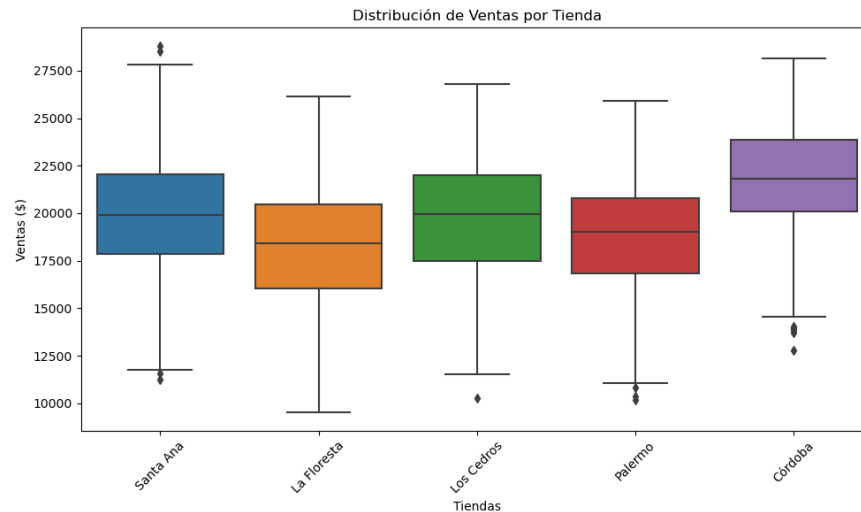


Figure 1: Diagrama de caja de las ventas de cada tienda

3.3 problema 3. Resultados del Análisis Estadístico

Tiendas Destacadas

La tienda con mayor promedio de ventas es	Córdoba (\$21,784.30)
La tienda con menor promedio de ventas es	La Floresta (\$18,049.11)

Prueba de Hipótesis

Se plantea la siguiente prueba de hipótesis:

H_0 : No hay diferencia significativa entre medias ($\mu_C - \mu_F = 0$).

H_1 : Hay diferencia significativa entre medias ($\mu_C - \mu_F \neq 0$).

Resultados de la Prueba ANOVA

La prueba se realizó empleando la función `ttest_ind` de la librería `scipy`.

- Estadístico t : 16.2497
- Valor p : 8.6900×10^{-51}

4 Conclusiones

- Problema 1.

El intervalo de confianza del 95% nos indica que podemos estar 95% seguros de que la verdadera media de ventas se encuentra dentro de ese rango. El intervalo de confianza del 99% es más amplio pero nos da mayor certeza, indicando que podemos estar 99% seguros de que la verdadera media está en ese rango.

Los intervalos más amplios indican mayor variabilidad en las ventas de ese mes. Los intervalos más estrechos indican ventas más consistentes.

Febrero mostró la mayor variabilidad en las ventas, lo que podría indicar mayor incertidumbre en las predicciones para este mes.

Esta información puede ser útil para la planificación de inventario y recursos, prestando especial atención a los meses con mayor variabilidad.

- Problema 2.

Dado que el valor p (2.3354×10^{-58}) es considerablemente menor que el nivel de significancia $\alpha = 0.05\%$, se rechaza la hipótesis nula (H_0). Esto implica que existe al menos una tienda cuyas ventas promedio son estadísticamente diferentes de las demás.

También se puede observar con el diagrama de cajas como las ventas de cada tienda se distribuyen de forma similar pero no idénticas, así como lo confirma el análisis ANOVA.

- Problema 3.

La diferencia es estadísticamente significativa ($p = 8.6900 \times 10^{-51}$). Por lo tanto, se rechaza la hipótesis nula (H_0) y se concluye que hay una diferencia significativa entre las medias de las ventas de las tiendas Córdoba y La Floresta.