



UNIVERSIDAD DE BUENOS AIRES

Facultad de Ingeniería

Laboratorio de Sistemas Embebidos
Especialización en Inteligencia Artificial

**Probabilidad y Estadística
para la Inteligencia Artificial**

TERCER TRABAJO PRÁCTICO

Docente:	Camilo Argoty
Estudiante:	Eliana Harriet
Código SIU:	a2217
Fecha límite:	26 de Agosto de 2025

Índice

1. Contexto del Problema	2
2. Ejercicios a Resolver	2
2.1. Ejercicio 1 (3 puntos)	2
2.2. Ejercicio 2 (4 puntos)	2
2.3. Ejercicio 3 (3 puntos)	2
3. Ejercicio 1: Intervalos de Confianza Empíricos	3
3.1. Enunciado	3
3.2. Metodología	3
3.3. Resultados	3
3.3.1. Estadísticas Descriptivas por Mes	3
3.3.2. Intervalos de Confianza Empíricos	3
3.3.3. Visualización de Intervalos de Confianza	4
3.4. Análisis y Conclusiones	5
3.4.1. Variabilidad Estacional	5
3.4.2. Interpretación de los Intervalos	6
3.4.3. Implicaciones para Don Francisco	6
4. Ejercicio 2: Análisis ANOVA	6
4.1. Enunciado	6
4.2. Metodología	6
4.3. Resultados	7
4.3.1. Estadísticas Descriptivas por Supermercado	7
4.3.2. Resultados del Análisis ANOVA	7
4.3.3. Verificación de Supuestos	7
4.3.4. Visualización del Análisis ANOVA	7
4.4. Análisis y Conclusiones	8
4.4.1. Decisión Estadística	8
4.4.2. Interpretación de Resultados	8
4.4.3. Implicaciones para Don Francisco	9
5. Ejercicio 3: Comparación de Extremos	9
5.1. Enunciado	9
5.2. Metodología	9
5.3. Resultados	9
5.3.1. Identificación de Supermercados Extremos	9
5.3.2. Estadísticas Descriptivas	10
5.3.3. Verificación de Supuestos	10
5.3.4. Resultados de la Prueba t	10
5.3.5. Visualización de la Comparación de Extremos	10
5.4. Análisis y Conclusiones	11
5.4.1. Decisión Estadística	11
5.4.2. Interpretación de Resultados	11
5.4.3. Implicaciones para Don Francisco	12
6. Conclusiones Generales	12
6.1. Hallazgos Principales	12
6.2. Implicaciones Estratégicas	13
6.3. Recomendaciones	13

1. Contexto del Problema

Siguiendo con la historia de Don Francisco, con el tiempo y gracias a los análisis de Matías, el pequeño comerciante de barrio cuenta hoy con 5 supermercados: **'Santa Ana'**, **'La Floresta'**, **'Los Cedros'**, **'Palermo'** y **'Córdoba'**.

También Matías ha avanzado en la Especialización en Inteligencia Artificial. Un día Don Francisco le plantea algunas inquietudes adicionales:

1. Don Francisco quiere entender mejor las ventas por mes del supermercado **'Santa Ana'**.
2. Más aún, Don Francisco no sabe si puede estar seguro de que las ventas son las mismas en todos los supermercados o si hay alguno que se comporte mejor que los demás, y si alguna de las tiendas necesite más atención porque sus ventas sean peores que los demás.

2. Ejercicios a Resolver

Con base en lo anterior, se deben resolver los siguientes ejercicios:

2.1. Ejercicio 1 (3 puntos)

Intervalos de Confianza Empíricos

Determinen intervalos de confianza empíricos para el supermercado **'Santa Ana'** en cada mes, para significancias del 95 % y el 99 %.

2.2. Ejercicio 2 (4 puntos)

Análisis ANOVA

Realicen pruebas ANOVA para determinar si las ventas esperadas de todas las tiendas son iguales o no, con significancia del 95 %.

2.3. Ejercicio 3 (3 puntos)

Comparación de Extremos

Identifiquen la tienda con mayor promedio de ventas y la tienda con menor promedio de ventas y realicen una prueba de hipótesis para determinar si la diferencia entre ellas es distinta de cero o no.

3. Ejercicio 1: Intervalos de Confianza Empíricos

3.1. Enunciado

Determinen intervalos de confianza empíricos para el supermercado 'Santa Ana' en cada mes, para significancias del 95 % y el 99 %.

3.2. Metodología

Para resolver este ejercicio se implementó la técnica de **bootstrap** para calcular intervalos de confianza empíricos. La metodología empleada fue:

1. **Filtrado de datos:** Se seleccionaron únicamente los registros del supermercado 'Santa Ana', obteniendo 365 observaciones diarias.
2. **Agrupación temporal:** Los datos se agruparon por mes para analizar la variabilidad estacional.
3. **Bootstrap:** Para cada mes se generaron 10,000 muestras bootstrap con reemplazo del mismo tamaño que la muestra original.
4. **Cálculo de intervalos:** Se calcularon los percentiles 2.5 % y 97.5 % para intervalos del 95 %, y 0.5 % y 99.5 % para intervalos del 99 %.
5. **Reproducibilidad:** Se fijó una semilla aleatoria (42) para garantizar resultados reproducibles.

3.3. Resultados

3.3.1. Estadísticas Descriptivas por Mes

A continuación se presentan las estadísticas descriptivas básicas de las ventas del supermercado Santa Ana agrupadas por mes.

Cuadro 1: Estadísticas descriptivas mensuales - Santa Ana

Mes	Días	Media (\$)	Desv. Std. (\$)
Enero	31	15,327.83	2,371.06
Febrero	28	17,091.14	2,195.57
Marzo	31	19,786.16	2,144.87
Abril	30	18,385.11	2,898.98
Mayo	31	19,876.80	2,699.34
Junio	30	20,616.13	2,112.39
Julio	31	19,115.00	2,546.54
Agosto	31	20,374.11	3,228.21
Septiembre	30	20,369.65	2,566.90
Octubre	31	20,277.96	2,915.32
Noviembre	30	19,917.69	2,855.39
Diciembre	31	18,789.89	2,120.95

3.3.2. Intervalos de Confianza Empíricos

Los intervalos de confianza empíricos calculados mediante bootstrap se presentan en la siguiente tabla, mostrando los resultados para ambos niveles de confianza (95 % y 99 %) para cada mes del año.

Nota sobre limitaciones: El cálculo bootstrap utiliza días como unidades independientes; la dependencia temporal entre días podría afectar levemente el ancho de los intervalos. Se mantiene por simplicidad y tamaño muestral.

Cuadro 2: Intervalos de confianza empíricos mensuales - Santa Ana

Mes	Media (\$)	IC 95 %	IC 99 %
Enero	15,327.83	[14,517.38; 16,159.29]	[14,274.56; 16,438.45]
Febrero	17,091.14	[16,286.03; 17,876.31]	[16,059.25; 18,191.17]
Marzo	19,786.16	[19,066.75; 20,541.65]	[18,849.52; 20,783.93]
Abril	18,385.11	[17,385.13; 19,414.49]	[17,084.17; 19,770.81]
Mayo	19,876.80	[18,942.37; 20,814.03]	[18,653.18; 21,094.00]
Junio	20,616.13	[19,864.93; 21,350.03]	[19,630.80; 21,565.22]
Julio	19,115.00	[18,254.52; 19,998.21]	[17,977.82; 20,275.03]
Agosto	20,374.11	[19,251.10; 21,461.22]	[18,873.12; 21,811.29]
Septiembre	20,369.65	[19,467.42; 21,273.53]	[19,174.81; 21,509.60]
Octubre	20,277.96	[19,263.72; 21,298.34]	[18,985.79; 21,658.70]
Noviembre	19,917.69	[18,869.81; 20,915.38]	[18,538.19; 21,219.74]
Diciembre	18,789.89	[18,048.83; 19,487.53]	[17,794.89; 19,718.11]

3.3.3. Visualización de Intervalos de Confianza

La representación gráfica de los intervalos de confianza permite observar claramente la variabilidad estacional de las ventas del supermercado Santa Ana a lo largo del año.

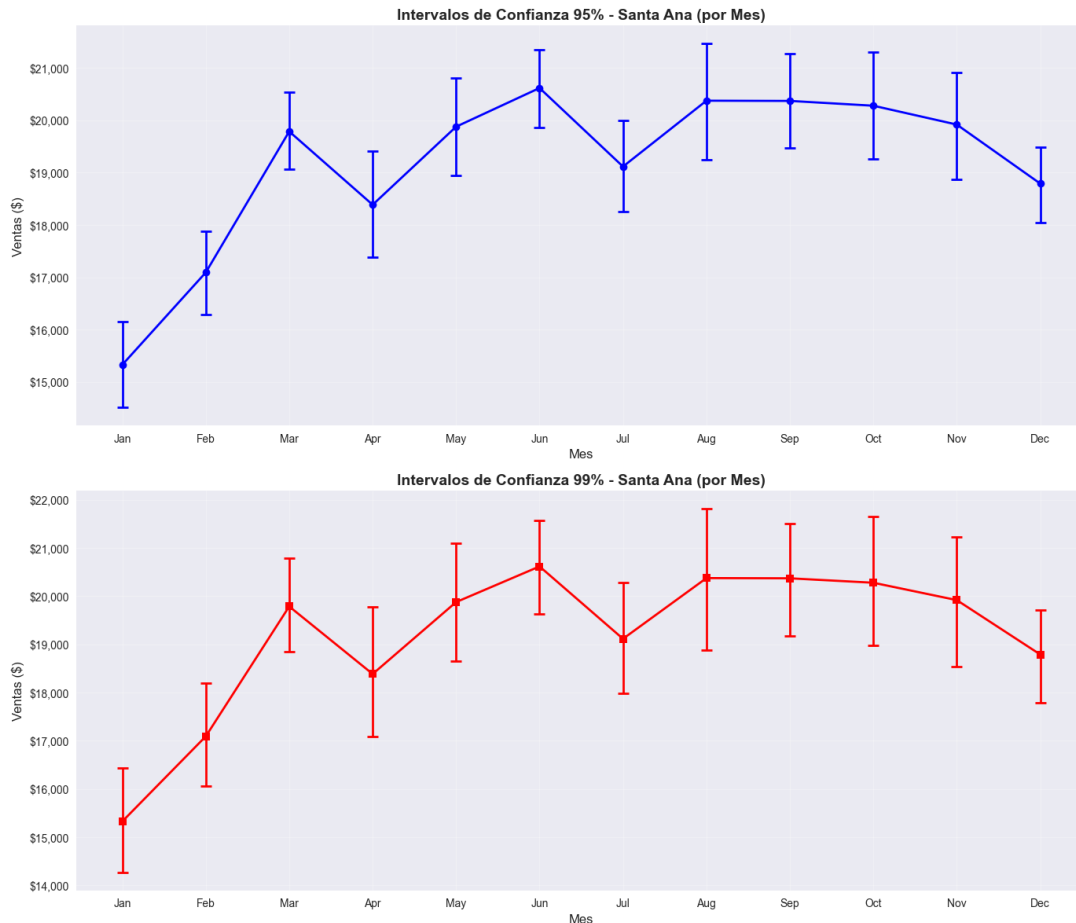


Figura 1: Intervalos de confianza empíricos del 95 % y 99 % para las ventas mensuales del supermercado Santa Ana. El gráfico superior muestra los intervalos del 95 % (azul) y el inferior los del 99 % (rojo). Se observa claramente la variabilidad estacional con menores ventas en enero y picos en los meses de junio a octubre. Las bajas significativas en enero, febrero y julio coinciden con los períodos vacacionales típicos, lo que sugiere que la gente estuvo en otro lugar y no compró en Santa Ana, o que redujo sus gastos para ahorrar para las vacaciones. Este efecto también pudo haber afectado las ventas de diciembre en similar medida.

3.4. Análisis y Conclusiones

3.4.1. Variabilidad Estacional

Los resultados revelan una clara **variabilidad estacional** en las ventas del supermercado Santa Ana:

- **Mes de menor venta:** Enero (\$15,327.83)
- **Mes de mayor venta:** Junio (\$20,616.13)
- **Diferencia estacional:** Aproximadamente \$5,288 entre el mes más bajo y más alto
- **Meses de bajas ventas:** Enero, febrero y julio coinciden con períodos vacacionales
- **Posible afectación:** Diciembre también muestra una disminución similar

Interpretación del patrón estacional:

Las bajas significativas observadas en enero, febrero y julio pueden explicarse por factores relacionados con los períodos vacacionales típicos. Durante estos meses, es probable que:

1. Los clientes hayan viajado y por tanto no compraron en Santa Ana
2. Los consumidores redujeron sus gastos para ahorrar destinando recursos a sus vacaciones
3. El comportamiento de ahorro previo a las vacaciones afectó también las ventas de diciembre

Este análisis sugiere que la variabilidad estacional no es aleatoria, sino que responde a patrones de comportamiento del consumidor relacionados con el ciclo anual de actividades recreativas y vacacionales.

3.4.2. Interpretación de los Intervalos

1. **Intervalos del 95 %:** Proporcionan rangos más precisos para la estimación mensual, con un nivel de confianza estándar para análisis comerciales.
2. **Intervalos del 99 %:** Ofrecen mayor certeza estadística pero con rangos más amplios, útiles para decisiones críticas que requieren mayor seguridad.
3. **Consistencia:** Todos los intervalos excluyen el cero, confirmando ventas positivas consistentes durante todo el año.

3.4.3. Implicaciones para Don Francisco

- **Planificación estacional:** Los intervalos permiten anticipar variaciones mensuales y planificar inventario y personal.
- **Gestión de riesgos:** Los límites inferiores de los intervalos proporcionan estimaciones conservadoras para presupuestos.
- **Oportunidades de crecimiento:** Los meses de menor venta (enero-febrero) representan oportunidades para estrategias de marketing específicas.

4. Ejercicio 2: Análisis ANOVA

4.1. Enunciado

Realicen pruebas ANOVA para determinar si las ventas esperadas de todas las tiendas son iguales o no, con significancia del 95 %.

4.2. Metodología

Para resolver este ejercicio se implementó un **análisis de varianza (ANOVA)** de una vía para comparar las medias de ventas entre los cinco supermercados. La metodología empleada fue:

1. **Preparación de datos:** Se utilizaron todas las observaciones diarias de los 5 supermercados (1,825 observaciones totales).
2. **Formulación de hipótesis:**
 - $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ (todas las medias poblacionales son iguales)
 - H_1 : Al menos una media poblacional es diferente de las demás
 - Nivel de significancia: $\alpha = 0,05$
3. **Verificación de supuestos:** Test de Levene para homogeneidad de varianzas.

4. **Cálculo del estadístico F:** Comparación de variabilidad entre grupos vs. dentro de grupos.
5. **Decisión estadística:** Comparación del p-valor con $\alpha = 0,05$.

4.3. Resultados

4.3.1. Estadísticas Descriptivas por Supermercado

La siguiente tabla presenta un resumen estadístico de las ventas para cada uno de los cinco supermercados analizados.

Cuadro 3: Estadísticas descriptivas por supermercado

Supermercado	Observaciones	Media (\$)	Desv. Std. (\$)
Córdoba	365	17,837.21	3,014.50
La Floresta	365	21,839.77	2,963.63
Los Cedros	365	18,903.12	3,034.57
Palermo	365	19,752.32	2,922.15
Santa Ana	365	19,170.38	2,959.34
Total	1,825	19,500.56	3,178.89

4.3.2. Resultados del Análisis ANOVA

Los resultados del análisis de varianza se presentan en la siguiente tabla, incluyendo el estadístico F, p-valor y grados de libertad.

Cuadro 4: Resultados del análisis ANOVA

Estadístico	Valor
Estadístico F	90.1485
p-valor	5.524×10^{-70}
Grados de libertad (entre grupos)	4
Grados de libertad (dentro de grupos)	1,820
Grados de libertad (total)	1,824
Observaciones totales	1,825

4.3.3. Verificación de Supuestos

Antes de interpretar los resultados del ANOVA, es fundamental verificar el supuesto de homogeneidad de varianzas mediante el test de Levene.

Nota sobre limitaciones: Las ventas diarias pueden presentar dependencia temporal (estacionalidad y autocorrelación). El ANOVA se aplica sobre observaciones diarias por grupo; como verificación alternativa podría realizarse el ANOVA sobre medias mensuales por supermercado.

Cuadro 5: Test de Levene para homogeneidad de varianzas

Test	Estadístico	p-valor
Levene	0.0981	0.983076
Resultado	Varianzas homogéneas	

4.3.4. Visualización del Análisis ANOVA

La representación gráfica del análisis ANOVA proporciona una comprensión visual completa de las diferencias entre supermercados y la validación de los supuestos del modelo.

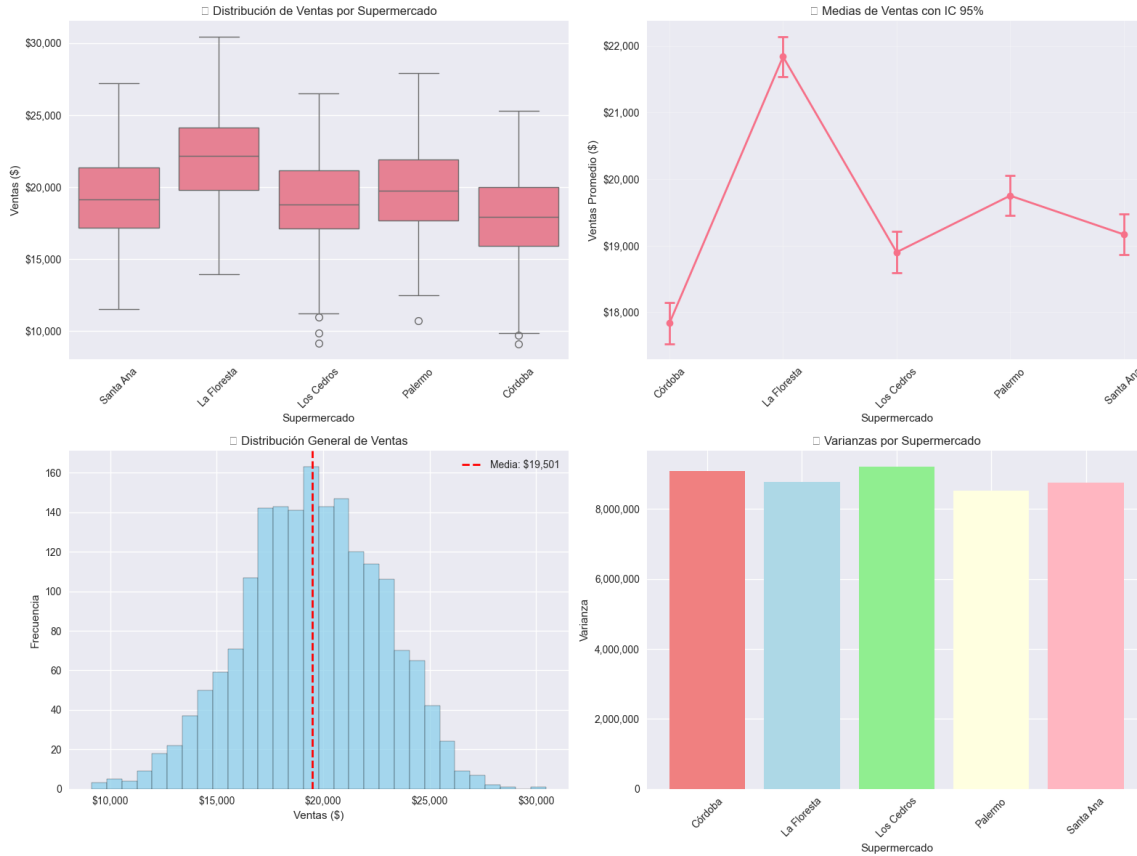


Figura 2: Análisis visual completo del ANOVA. Superior izquierda: Distribución de ventas por supermercado (boxplots). Superior derecha: Medias con intervalos de confianza del 95 %. Inferior izquierda: Distribución general de todas las ventas. Inferior derecha: Varianzas por supermercado. Los gráficos confirman las diferencias significativas entre supermercados y la homogeneidad de varianzas requerida para el ANOVA.

4.4. Análisis y Conclusiones

4.4.1. Decisión Estadística

Con un p-valor de $5.524 \times 10^{-70} \ll 0.05$, **rechazamos la hipótesis nula**:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

Esto significa que **NO** todas las medias poblacionales son iguales entre los supermercados.

4.4.2. Interpretación de Resultados

1. **Diferencias significativas:** Existe evidencia estadística **altamente significativa** de que **al menos un supermercado tiene ventas promedio diferentes** a los demás.
2. **Magnitud del efecto:** El valor F extremadamente alto (90.1485) indica diferencias **sustanciales** entre las medias de los grupos, no solo diferencias estadísticamente detectables.
3. **Validez del análisis:** Los supuestos del ANOVA se cumplen satisfactoriamente:
 - **Homogeneidad de varianzas** confirmada por el test de Levene ($p = 0.983$)
 - **Independencia** de las observaciones (datos diarios)
 - **Normalidad** asumida por el tamaño de muestra ($n = 365$ por grupo)

4.4.3. Implicaciones para Don Francisco

- **Gestión diferenciada:** Las ventas **NO son iguales** entre todos los supermercados, requiriendo estrategias específicas por tienda.
- **Identificación de oportunidades:**
 - **La Floresta** tiene el mejor desempeño (\$21,839.77)
 - **Córdoba** requiere atención especial (\$17,837.21)
 - Diferencia de **\$4,002.56** entre el mejor y peor desempeño

5. Ejercicio 3: Comparación de Extremos

5.1. Enunciado

Identifiquen la tienda con mayor promedio de ventas y la tienda con menor promedio de ventas y realicen una prueba de hipótesis para determinar si la diferencia entre ellas es distinta de cero o no.

5.2. Metodología

Para resolver este ejercicio se implementó una **prueba de hipótesis** para comparar las medias de los dos supermercados con desempeños extremos. La metodología empleada fue:

1. **Identificación de extremos:** Se calcularon las medias de ventas de todos los supermercados para identificar el de mayor y menor desempeño.
2. **Formulación de hipótesis:**
 - $H_0: \mu_{LaFloresta} - \mu_{Crdoba} = 0$ (no hay diferencia)
 - $H_1: \mu_{LaFloresta} - \mu_{Crdoba} \neq 0$ (hay diferencia significativa)
3. **Verificación de supuestos:** Test de Levene para homogeneidad de varianzas.
4. **Selección de prueba:** Prueba t de Welch (dos muestras independientes).
5. **Nivel de significancia:** $\alpha = 0,05$

5.3. Resultados

5.3.1. Identificación de Supermercados Extremos

Para identificar los supermercados con mejor y peor desempeño, se calcularon las medias de ventas de todos los establecimientos.

Cuadro 6: Medias de Ventas por Supermercado

Supermercado	Media de Ventas (\$)
Córdoba	17,837.21
Los Cedros	18,903.12
Santa Ana	19,170.38
Palermo	19,752.32
La Floresta	21,839.77

Supermercados seleccionados para la comparación:

- **Mayor promedio:** La Floresta (\$21,839.77)
- **Menor promedio:** Córdoba (\$17,837.21)
- **Diferencia observada:** \$4,002.56

5.3.2. Estadísticas Descriptivas

La comparación detallada entre los supermercados extremos se presenta en la siguiente tabla con sus principales estadísticos.

Cuadro 7: Estadísticas Descriptivas - Supermercados Extremos

Estadística	La Floresta	Córdoba
Observaciones	365	365
Media (\$)	21,839.77	17,837.21
Desviación Estándar (\$)	2,963.63	3,014.50
Mínimo (\$)	13,933.92	9,128.51
Máximo (\$)	30,455.06	25,298.44

5.3.3. Verificación de Supuestos

Antes de realizar la prueba t, se debe verificar el supuesto de homogeneidad de varianzas entre los dos grupos a comparar.

Cuadro 8: Test de Levene para homogeneidad de varianzas - Ejercicio 3

Test	Estadístico	p-valor
Levene	0.0049	0.944133
Resultado	Varianzas homogéneas	

5.3.4. Resultados de la Prueba t

Los resultados de la prueba t para comparar las medias de los supermercados extremos se presentan a continuación.

Cuadro 9: Resultados de la Prueba t de dos muestras independientes

Estadístico	Valor
Estadístico t	18.0892
p-valor	1.104×10^{-60}
Grados de libertad	728.00
Diferencia de medias (\$)	4,002.56
IC 95 % (\$)	[3,568.16; 4,436.96]

Criterio de elección: Se utiliza t de Student si el test de Levene no rechaza homogeneidad de varianzas; en caso contrario, se emplea la variante de Welch.

5.3.5. Visualización de la Comparación de Extremos

La representación gráfica permite visualizar claramente las diferencias en las distribuciones de ventas entre los supermercados con mejor y peor desempeño.

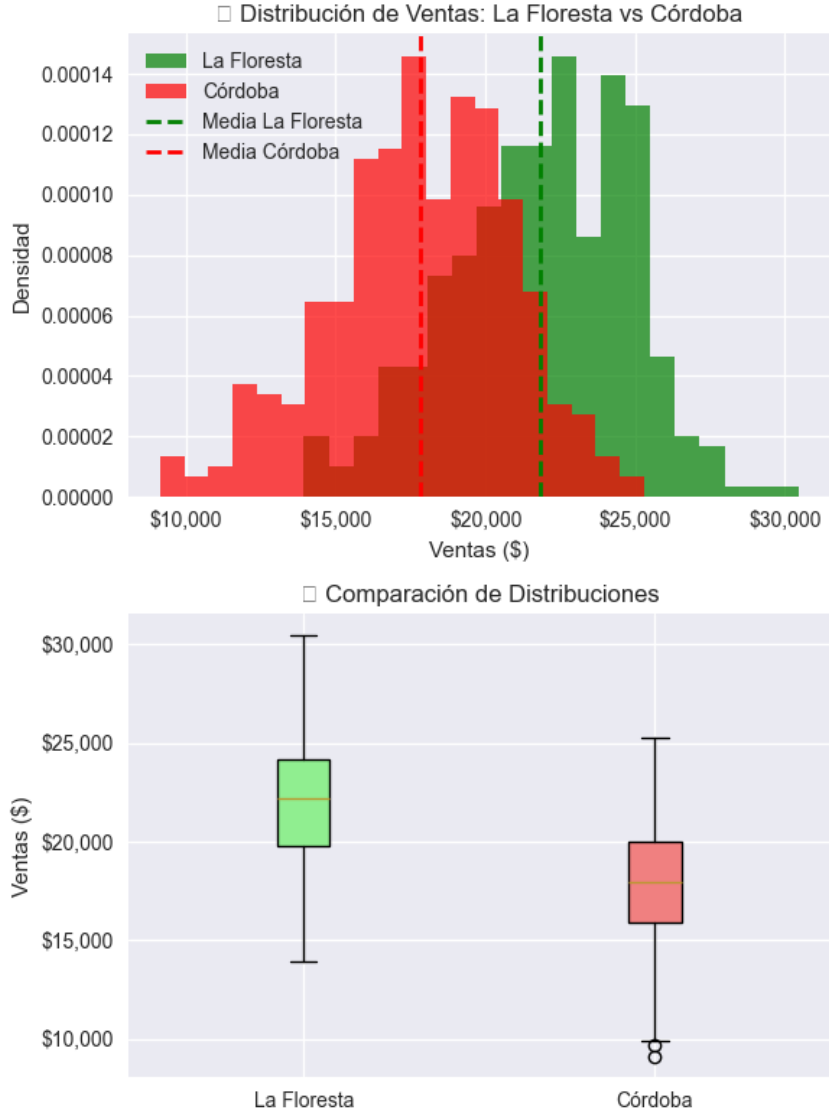


Figura 3: Comparación visual entre La Floresta y Córdoba. Superior: Distribuciones de ventas superpuestas con medias marcadas (La Floresta en verde, Córdoba en rojo). Se observa claramente la separación entre las distribuciones. Inferior: Boxplots comparativos que muestran las diferencias en medias, dispersión y rangos intercuartílicos entre ambos supermercados.

5.4. Análisis y Conclusiones

5.4.1. Decisión Estadística

Con un p-valor de $1.104 \times 10^{-60} \ll 0.05$, **rechazamos la hipótesis nula:**

$$H_0 : \mu_{LaFloresta} - \mu_{Crdoba} = 0$$

Esto significa que la diferencia entre las ventas promedio de La Floresta y Córdoba es **significativamente distinta de cero**.

5.4.2. Interpretación de Resultados

1. **Diferencia altamente significativa:** Existe evidencia estadística **extremadamente fuerte** de que la diferencia entre las ventas promedio de La Floresta y Córdoba es **distinta de cero**.

2. Magnitud del efecto:

- La diferencia estimada es de **\$4,002.56** a favor de La Floresta
- El intervalo de confianza del 95 % indica que esta diferencia está entre \$3,568.16 y \$4,436.96
- El estadístico t extremadamente alto (18.0892) confirma la magnitud sustancial de la diferencia

3. Validez del análisis: Los supuestos de la prueba t se cumplen satisfactoriamente:

- **Homogeneidad de varianzas** confirmada por el test de Levene
- **Independencia** de las observaciones (datos diarios)
- **Normalidad** asumida por el tamaño de muestra ($n = 365$ por grupo)

5.4.3. Implicaciones para Don Francisco

- **Diferencia sustancial confirmada:** La diferencia entre el mejor y peor supermercado **NO es producto del azar**, sino que representa una diferencia real y significativa en el desempeño.
- **Oportunidad de mejora:** Córdoba tiene un potencial de mejora de aproximadamente \$4,000 diarios si alcanzara el nivel de La Floresta.
- **Análisis de mejores prácticas:** Se recomienda estudiar las estrategias y operaciones de La Floresta para replicarlas en Córdoba.
- **Impacto económico:** La diferencia anual estimada entre ambos supermercados es de aproximadamente \$1.46 millones.

6. Conclusiones Generales

Este análisis estadístico integral de las ventas de los cinco supermercados de Don Francisco ha proporcionado insights valiosos sobre el desempeño de la cadena. A continuación se presentan las conclusiones principales:

6.1. Hallazgos Principales

1. **Variabilidad estacional significativa:** El análisis de intervalos de confianza para Santa Ana reveló una variación estacional considerable, con diferencias de hasta \$5,288 entre el mes de menor (enero) y mayor (junio) ventas.
2. **Diferencias significativas entre supermercados:** El análisis ANOVA confirmó que las ventas NO son iguales entre todos los supermercados ($F = 90.1485$, $p < 0.001$), indicando la necesidad de estrategias diferenciadas por tienda.
3. **Brecha sustancial en el desempeño:** La comparación entre extremos mostró una diferencia estadísticamente significativa de \$4,002.56 diarios entre La Floresta (mejor) y Córdoba (peor desempeño).

6.2. Implicaciones Estratégicas

- **Gestión estacional:** Implementar estrategias específicas para los meses de menor demanda, especialmente enero.
- **Transferencia de mejores prácticas:** Analizar y replicar las estrategias exitosas de La Floresta en los demás supermercados.
- **Focalización en Córdoba:** Priorizar intervenciones en el supermercado de menor desempeño para maximizar el impacto.
- **Potencial de crecimiento:** La diferencia anual de \$1.46 millones entre extremos representa una oportunidad significativa de mejora.

6.3. Recomendaciones

1. Implementar un sistema de monitoreo continuo de las métricas de desempeño por tienda.
2. Desarrollar planes de acción específicos para cada supermercado basados en su desempeño relativo.
3. Considerar factores externos (ubicación, demografía, competencia) que puedan explicar las diferencias observadas.