

CLASE

8

ESTADÍSTICA II

Pruebas no paramétricas II

Modelos aplicados

INTRO DUCCIÓN

DE LA CLASE

GRADO / POSGRADO / TECNOLOGÍAS



En esta octava semana se profundizará en las pruebas no paramétricas como herramientas fundamentales para la comparación de grupos cuando los supuestos de los métodos tradicionales no se cumplen, abordando técnicas reconocidas como Mann-Whitney y Kruskal-Wallis, junto con la evaluación de los supuestos necesarios para su correcta aplicación y la interpretación de los resultados, de manera que los estudiantes puedan desarrollar competencias sólidas en el análisis de datos ordinales o no normales, comprender las ventajas y limitaciones de estos métodos y aplicar los resultados de manera rigurosa en la toma de decisiones dentro de contextos administrativos, investigativos y profesionales, promoviendo un enfoque estadístico flexible y crítico que permita extraer conclusiones fundamentadas y pertinentes a partir de información compleja y diversa.

De manera complementaria, se explorarán modelos aplicados avanzados, incluyendo series de tiempo para el análisis de tendencias y estacionalidad y la regresión logística para el estudio de variables binarias, enfatizando la interpretación de resultados y la aplicación práctica de estos modelos en escenarios reales, de modo que los estudiantes puedan consolidar habilidades para proyectar comportamientos futuros, generar predicciones confiables y diseñar soluciones basadas en datos, integrando de manera coherente la teoría estadística, el análisis empírico y la toma de decisiones estratégicas, fortaleciendo así la capacidad para abordar problemas complejos en entornos académicos, corporativos y de investigación con un enfoque analítico riguroso y profesional.

Educación de calidad



INTRO DUCCIÓN CIÓN DE LA CLASE

GRADO / POSGRADO / TECNOLOGÍAS

Estudia a tu tiempo
son interrupciones

Resultado o resultados de aprendizaje
que será abordado con el contenido
de la clase.

- 1) Identificar los conceptos, definiciones, propiedades y teoremas de la Estadística Inferencial como base del análisis de datos en el contexto de su área de conocimiento.
- 2) Analizar las técnicas de muestreo, estimadores y los contrastes de hipótesis, en el contexto de su área del conocimiento.
- 3) Interpretar los resultados de la aplicación de técnicas de la Estadística Inferencial en el contexto de su área de conocimiento.

15. Pruebas no paramétricas II

Esta sección amplía el estudio de las pruebas no paramétricas enfocándose en la comparación de grupos cuando los datos no cumplen los supuestos paramétricos. Se destacan métodos como Mann-Whitney y Kruskal-Wallis, explicando su utilidad para evaluar diferencias entre dos o más grupos de manera robusta, además de considerar los supuestos necesarios para garantizar la validez de los resultados.

Además, se enfatiza que estas pruebas permiten obtener conclusiones fiables sin depender de la distribución normal de los datos, siendo herramientas especialmente útiles en investigaciones con muestras pequeñas, escalas ordinales o presencia de datos atípicos.

15.1 Comparación de grupos

Las pruebas no paramétricas para comparación de grupos se basan en rangos en lugar de valores absolutos, lo que las hace robustas ante distribuciones no normales y outliers.

Se estudian las pruebas Mann-Whitney y Kruskal-Wallis, detallando cómo estas técnicas evalúan diferencias significativas entre grupos y cuándo su aplicación es apropiada según la estructura de los datos. La comprensión de los supuestos y limitaciones de cada prueba es esencial para garantizar que los resultados sean interpretables y relevantes para los objetivos del estudio.

Asimismo, se analiza la interpretación de los resultados, destacando la importancia de contextualizar las conclusiones y considerar la magnitud de las diferencias, lo que fortalece la capacidad de tomar decisiones basadas en evidencia cuantitativa confiable.

15.1.1 Mann-Whitney

La **prueba Mann-Whitney U** (también conocida como prueba de suma de rangos de Wilcoxon) se utiliza para comparar dos grupos independientes cuando los datos no cumplen los supuestos de la prueba t.

Hipótesis:

- H_0 : Las distribuciones de los dos grupos son iguales (no hay diferencia en la mediana)
- H_1 : Las distribuciones difieren (existe diferencia en la mediana)

Se calcula ordenando todos los datos de ambos grupos y comparando las sumas de rangos. Es especialmente útil con datos ordinales o continuos no normales.

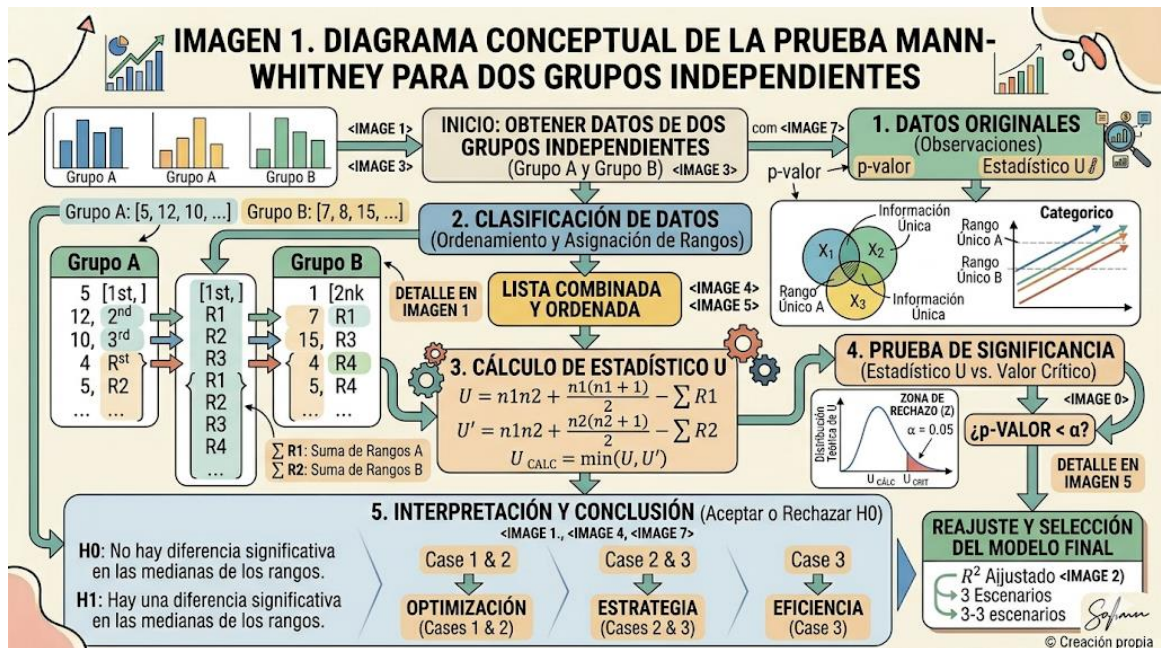


Imagen 1. Diagrama conceptual de la prueba Mann-Whitney para dos grupos independientes (creación propia del autor).

En un contexto empresarial, la prueba Mann-Whitney permite comparar grupos de manera robusta cuando los datos no siguen una distribución normal. Por ejemplo, se puede evaluar si dos regiones presentan diferencias en satisfacción del cliente medida en escalas ordinales, o comparar la eficiencia de dos métodos de entrega cuando los tiempos no son simétricos. Al centrarse en las medianas y no en las medias, esta prueba proporciona una visión más realista de diferencias centrales, evitando que valores extremos influyan indebidamente en las conclusiones. Esto facilita decisiones estratégicas basadas en evidencia, como ajustar procesos regionales, mejorar productos o rediseñar servicios según los resultados observados.

15.1.2 Kruskal-Wallis

La **prueba Kruskal-Wallis** es la versión no paramétrica del ANOVA de un factor. Se utiliza para comparar tres o más grupos independientes.

Hipótesis:

- H_0 : Todas las poblaciones tienen la misma distribución (medianas iguales)
- H_1 : Al menos una población difiere de las demás

Se basa en rangos de todos los datos combinados y sigue una distribución aproximada a Chi-cuadrado.

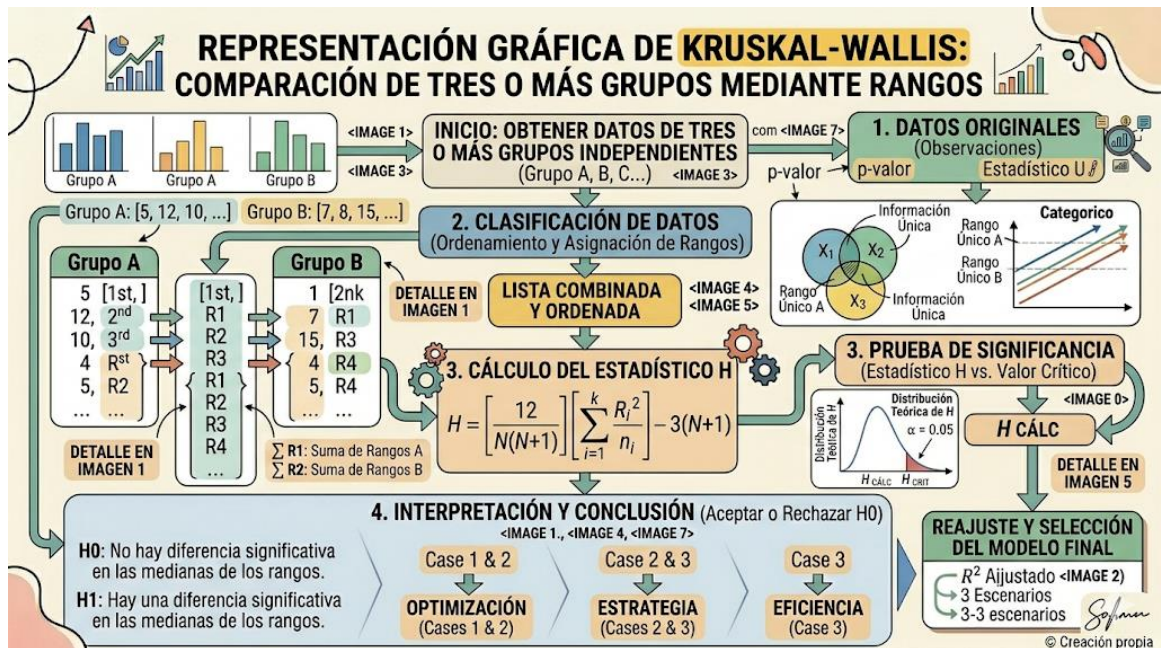


Imagen 2. Representación gráfica de Kruskal-Wallis: comparación de tres o más grupos mediante rangos (creación propia del autor).

Tanto Mann-Whitney como Kruskal-Wallis trabajan con rangos en lugar de valores crudos, lo que las hace menos sensibles a outliers y distribuciones asimétricas, resultandos ideales para datos reales de encuestas o mediciones empresariales.

15.1.3 Supuestos

Aunque son más flexibles que las pruebas paramétricas, estas pruebas requieren ciertos supuestos:

- Las observaciones son independientes.
- La variable dependiente es al menos ordinal (puede ser continua).
- Las distribuciones de los grupos tienen formas similares (para interpretar diferencias en medianas; si las formas difieren mucho, se interpreta diferencia en distribuciones generales).

No requieren normalidad ni homocedasticidad, lo que representa su principal ventaja.

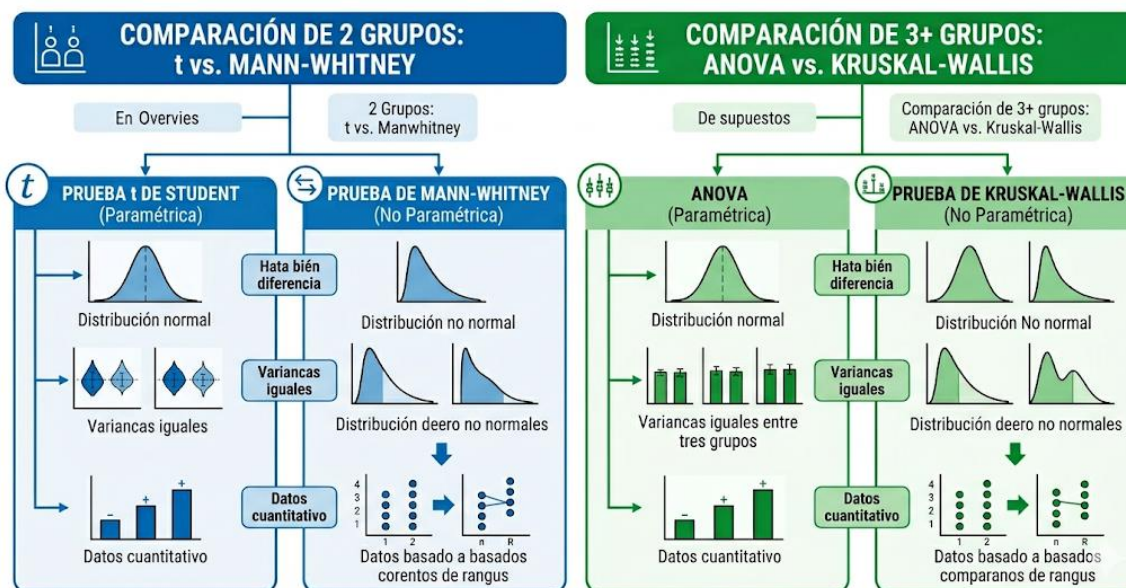


Imagen 3. Comparación visual de supuestos: prueba t vs. Mann-Whitney y ANOVA vs. Kruskal-Wallis (creación propia del autor).

En la práctica empresarial, la flexibilidad de Mann-Whitney para trabajar con datos no normales o con varianzas desiguales la hace muy útil en estudios de satisfacción del cliente, evaluaciones de desempeño o análisis de encuestas ordinales. Por ejemplo, al comparar la percepción de calidad de dos líneas de productos, esta prueba permite obtener conclusiones fiables incluso si las calificaciones no se distribuyen normalmente o si los tamaños de muestra difieren entre grupos. Además, su énfasis en la independencia de observaciones y la necesidad de escalas ordinales asegura que los resultados reflejen diferencias reales entre grupos, proporcionando información robusta para decisiones estratégicas y mejoras operativas.

15.2 Aplicación

Se enfatiza la aplicación práctica de las pruebas no paramétricas para comparar grupos en distintos contextos, evaluando ventajas como su flexibilidad y menor dependencia de supuestos estrictos, así como limitaciones relacionadas con la potencia estadística y la interpretación de resultados.

Además, se presentan ejemplos de uso en investigación y entornos profesionales, mostrando cómo estas pruebas permiten analizar datos complejos de manera efectiva y respaldar decisiones estratégicas basadas en hallazgos sólidos y confiables.

En entornos empresariales, las pruebas no paramétricas como Mann-Whitney o Kruskal-Wallis son especialmente valiosas cuando los datos no cumplen los supuestos de normalidad o cuando se trabaja con escalas ordinales, encuestas de satisfacción, calificaciones de desempeño o rankings de productos. Por ejemplo, permiten comparar la preferencia de clientes entre distintas versiones de un producto, evaluar la efectividad de dos métodos de capacitación o analizar diferencias en tiempos de entrega entre centros logísticos. Su flexibilidad garantiza que las decisiones estratégicas como ajustar procesos, diseñar incentivos o priorizar inversiones, se basen en evidencia sólida, incluso cuando los datos presentan distribuciones irregulares o tamaños de muestra desiguales.

15.2.1 Interpretación

La interpretación se centra en el valor p:

- Si $p < 0.05$, se rechaza $H_0 \rightarrow$ existe diferencia significativa entre los grupos.
- Para Mann-Whitney se interpreta la dirección con las medianas o sumas de rangos.
- Para Kruskal-Wallis, si es significativa, se deben realizar pruebas post-hoc no paramétricas (como Dunn o Bonferroni ajustado) para identificar qué grupos difieren.

Es importante reportar las medianas de cada grupo y el tamaño del efecto (por ejemplo, r de correlación de rangos).

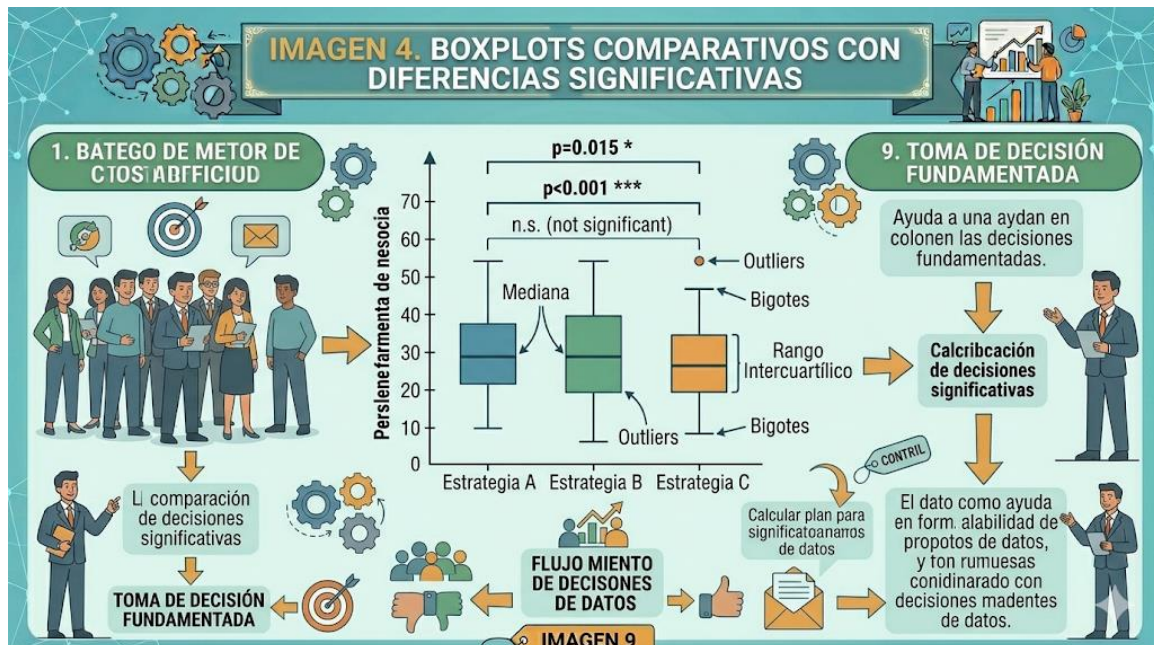


Imagen 4. Gráfico de boxplots comparativos con anotaciones de diferencias significativas (creación propia del autor).

15.2.2 Ventajas

Las principales ventajas de estas pruebas son:

- No requieren normalidad de los datos.
- Son robustas ante outliers y distribuciones asimétricas.
- Se pueden aplicar a datos ordinales (escalas Likert, rankings, etc.).
- Fáciles de interpretar en contextos donde los datos no son perfectos.
- Menor riesgo de conclusiones erróneas cuando los supuestos paramétricos fallan.

La gran ventaja de Mann-Whitney y Kruskal-Wallis es su robustez: permiten obtener resultados confiables incluso cuando los datos son sesgados, tienen outliers o provienen de escalas subjetivas como encuestas de satisfacción.

15.2.3 Limitaciones

A pesar de sus ventajas, tienen limitaciones:

- Menor poder estadístico que las pruebas paramétricas cuando los datos sí cumplen los supuestos de normalidad.
- No estiman parámetros poblacionales como medias (trabajan con medianas o distribuciones).

- La interpretación post-hoc en Kruskal-Wallis es menos directa que en ANOVA.
- No permiten controlar variables adicionales fácilmente (para eso se necesitarían modelos más avanzados).



Imagen 5. Infografía comparativa: ventajas y limitaciones de pruebas no paramétricas vs. paramétricas (creación propia del autor).

Enlace 1: Tutorial práctico sobre cómo realizar Mann-Whitney y Kruskal-Wallis en Excel o SPSS, incluyendo interpretación de resultados y pruebas post-hoc. [Ver en YouTube: Cómo hacer Pruebas No Paramétricas en Excel](#) Descripción del enlace: Ideal para aplicar inmediatamente estas pruebas en datos empresariales reales como encuestas o mediciones operativas.

16. Modelos aplicados

Esta sección aborda modelos estadísticos aplicados en distintos contextos, destacando su relevancia para la predicción, el análisis de tendencias y la toma de decisiones basadas en datos. Se estudian series de tiempo y regresión logística como herramientas fundamentales para analizar patrones, evaluar relaciones y proyectar resultados en escenarios reales.

Además, se enfatiza la comprensión de los elementos clave de cada modelo, incluyendo la identificación de tendencias y estacionalidad en series de tiempo, así como la interpretación de variables binarias en regresión logística, asegurando que los análisis sean precisos, significativos y aplicables a problemas prácticos.

16.1 Series de tiempo

Las series de tiempo son secuencias de datos ordenados cronológicamente. Su análisis permite identificar patrones, pronosticar valores futuros y apoyar la planificación empresarial.

Se estudian los componentes de las series de tiempo, incluyendo tendencia, estacionalidad y análisis de patrones, permitiendo identificar cambios a lo largo del tiempo y anticipar comportamientos futuros. La comprensión de estos elementos es esencial para planificar, pronosticar y tomar decisiones fundamentadas en datos históricos.

Asimismo, se analiza cómo aplicar técnicas de series de tiempo para evaluar patrones recurrentes y generar proyecciones confiables, fortaleciendo la capacidad de interpretar resultados y respaldar estrategias en contextos académicos, empresariales y profesionales.

16.1.1 Tendencia

La tendencia representa el movimiento general a largo plazo de la serie (crecimiento, decrecimiento o estabilidad). Puede ser lineal, exponencial o polinómica.

Ejemplos:

- Crecimiento sostenido de ventas online.
- Disminución gradual de costos de producción por mejoras tecnológicas.

Se puede estimar mediante regresión lineal sobre el tiempo o promedios móviles.

En el ámbito empresarial, identificar la tendencia de una serie de tiempo permite planificar estratégicamente recursos y operaciones. Por ejemplo, detectar un crecimiento sostenido en ventas online ayuda a dimensionar inventarios, personal y campañas de marketing; mientras que observar una disminución gradual de costos de producción puede orientar decisiones de inversión en tecnología o procesos. La estimación de la tendencia mediante regresión o promedios móviles proporciona un panorama claro de hacia dónde se dirige el negocio, facilitando decisiones basadas en datos históricos y anticipando posibles escenarios futuros.



Imagen 6. Serie de tiempo con tendencia lineal ascendente en ventas mensuales (creación propia del autor).

16.1.2 Estacionalidad

La estacionalidad son patrones repetitivos que ocurren en intervalos fijos (diarios, semanales, mensuales o anuales). Ejemplos comunes en negocios:

- Aumento de ventas en diciembre (Navidad).
- Pico de demanda de helados en verano.
- Mayor ausentismo laboral los lunes.

Se detecta mediante gráficos de descomposición o análisis de autocorrelación.

Identificar correctamente la tendencia y la estacionalidad permite a las empresas anticipar demanda, ajustar inventarios y planificar campañas, evitando tanto faltantes como excesos de stock.

16.1.3 Análisis

El análisis de series de tiempo combina:

- Descomposición (tendencia + estacionalidad + ciclo + irregular).
- Suavizado (promedios móviles, exponencial).
- Modelos avanzados (ARIMA, Holt-Winters) para pronóstico.

En la práctica empresarial se usa para:

- Pronosticar ventas futuras.
- Planificar producción y logística.
- Detectar anomalías (caídas inesperadas en ventas).



Imagen 7. Ejemplo de pronóstico con modelo de series de tiempo (creación propia del autor).

El análisis de series de tiempo no solo permite prever valores futuros, sino que también sirve como herramienta de diagnóstico y optimización. Por ejemplo, al identificar patrones estacionales, una empresa puede ajustar inventarios antes de picos de demanda, optimizar la programación de personal y planificar campañas de marketing. Asimismo, detectar ciclos económicos o fluctuaciones periódicas ayuda a anticipar necesidades financieras y ajustar estrategias de inversión. La combinación de descomposición, suavizado y modelos avanzados proporciona un marco sólido para decisiones basadas en datos, reduciendo riesgos y mejorando la eficiencia operativa en entornos dinámicos.

16.2 Regresión logística

La regresión logística se utiliza cuando la variable dependiente es binaria (0/1, sí/no, éxito/fracaso). A diferencia de la regresión lineal, predice probabilidades entre 0 y 1 mediante la función logística.

La ecuación básica es:

$$P(Y = 1) = \frac{1}{1 + e^{-(b_0 + b_1X_1 + \dots + b_kX_k)}}$$

En el contexto empresarial, la regresión logística es fundamental para predecir eventos binarios y tomar decisiones estratégicas basadas en probabilidades. Por ejemplo, se puede usar para estimar la probabilidad de que un cliente compre un producto, se dé de baja de un servicio o responda a una campaña de marketing. También es útil en riesgos financieros, como prever la probabilidad de impago de un crédito. La interpretación de los coeficientes permite identificar los factores que aumentan o disminuyen la probabilidad del evento, lo que facilita diseñar acciones correctivas o preventivas, optimizando recursos y mejorando la toma de decisiones basada en datos.

16.2.1 Variables binarias

La variable dependiente es dicotómica:

- ¿El cliente comprará? (1 = sí, 0 = no)
- ¿El cliente incumplirá el pago? (1 = default, 0 = no)
- ¿El empleado se quedará en la empresa el próximo año? (1 = retiene, 0 = renuncia)

Las variables independientes pueden ser continuas o categóricas (incluyendo dummies).

En la práctica, la correcta codificación de la variable dependiente y de las independientes es clave para la efectividad del modelo. Por ejemplo, transformar características categóricas en variables dummy permite evaluar cómo diferentes segmentos de clientes afectan la probabilidad de compra. Asimismo, incluir variables continuas como edad, ingresos o historial de compras permite afinar las predicciones. Esta combinación de variables dicotómicas y continuas facilita identificar patrones de comportamiento, priorizar clientes de alto valor, diseñar estrategias de retención o segmentar riesgos crediticios, proporcionando decisiones empresariales basadas en evidencia cuantitativa.

16.2.2 Interpretación

Los coeficientes se interpretan en términos de odds ratio (razón de probabilidades):

- Un coeficiente positivo aumenta la probabilidad del evento.
- $\text{Exp}(b) > 1$ indica que al aumentar la variable en una unidad, las probabilidades del evento se multiplican por $\text{exp}(b)$.

Se evalúa significancia con valores p y se usa el pseudo R^2 (McFadden) para medir el ajuste.

En regresión logística no interpretamos el coeficiente como cambio en Y , sino como cambio en la probabilidad o en las odds del evento. Esto permite tomar decisiones como aprobar o rechazar un crédito según el perfil del cliente.

16.2.3 Aplicaciones

La regresión logística tiene amplias aplicaciones empresariales:

- **Marketing:** Predecir probabilidad de compra o respuesta a una campaña.
- **Finanzas:** Evaluación de riesgo crediticio y detección de fraude.
- **Recursos Humanos:** Predecir rotación de personal (churn de empleados).
- **Operaciones:** Probabilidad de que un proveedor entregue a tiempo.
- **Salud ocupacional:** Probabilidad de accidente según condiciones laborales.

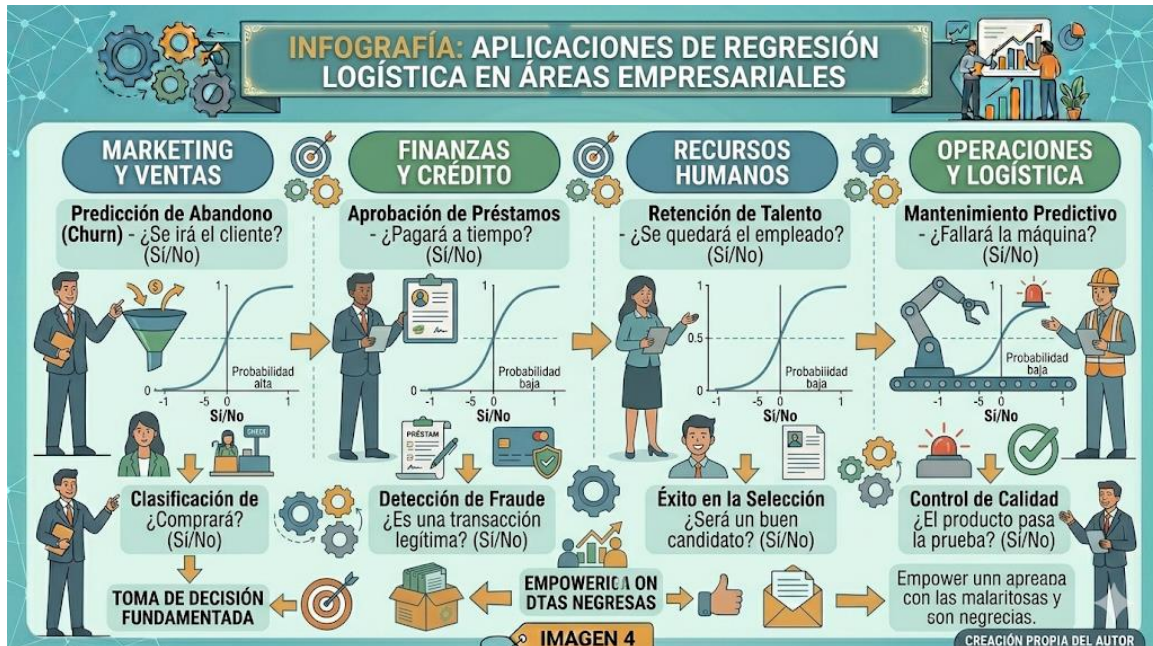


Imagen 8. Infografía con aplicaciones de regresión logística en áreas empresariales (creación propia del autor).

Enlace 2: Este video explica de forma clara el análisis de series de tiempo, tendencia y estacionalidad con ejemplos prácticos. [Ver en YouTube: Análisis de la tendencia en series temporales](#) **Descripción del enlace:** Este recurso visual te ayudará a descomponer series de tiempo y entender su aplicación en pronósticos empresariales.

RE FE REN CIAS

- Field, A. (2018). Discovering statistics using IBM SPSS statistics (5th ed.). SAGE Publications.
- Conover, W. J. (1999). Practical nonparametric statistics (3rd ed.). Wiley.
- Chatfield, C. (2003). The analysis of time series: An introduction (6th ed.). CRC Press.
- Agresti, A. (2018). An introduction to categorical data analysis (3rd ed.). Wiley.
- Montgomery, D. C., & Runger, G. C. (2018). Applied statistics and probability for engineers (7th ed.). Wiley.

Definición de los términos citados en la semana 8

- **Prueba Mann-Whitney:** Prueba no paramétrica que compara dos grupos independientes mediante rangos para determinar si sus distribuciones difieren significativamente, ideal cuando los datos no son normales.
- **Prueba Kruskal-Wallis:** Prueba no paramétrica equivalente al ANOVA de un factor que compara tres o más grupos independientes basándose en rangos de los datos.

Recursos de Profundización

1. Comparación de Grupos (Mann-Whitney, Kruskal-Wallis)

Descripción del enlace: Permite comparar grupos sin asumir normalidad mediante pruebas no paramétricas.

Autor: Carlos Julio Mayorga Arias

Enlace: [RECURSO 1](#)

2. Introducción a Series de Tiempo

Descripción del enlace: Simulador que permite analizar tendencia y estacionalidad en datos temporales.

Autor: Carlos Julio Mayorga Arias



Enlace: [RECURSO 2](#)

3. Regresión Logística Aplicada

Descripción del enlace: Permite modelar variables binarias y analizar probabilidades en contextos reales.

Autor: Carlos Julio Mayorga Arias

Enlace: [RECURSO 3](#)



síguenos en:



www.pucevirtual.puce.edu.ec