



UNIVERSIDAD AUTÓNOMA DE
NUEVO LEÓN



Facultad de Ciencias Físico y Matemáticas

Maestría en Ciencia de Datos

Métodos Estadísticos Multivariados

Proyecto
Análisis de vendedores

Grupo 002
Iván Gabriel Salinas Castillo 1856735
Álvaro Salgado López 2224885

MET Alejandra Cerda

Fecha de entrega 23/07/2024

Tabla de contenido

<i>I. INTRODUCCIÓN</i>	3
<i>II. DESARROLLO</i>	4
A. Análisis exploratorio de los datos	4
B. Pruebas de bondad de ajuste.....	6
C. Planteamiento de comparativas:	9
D. Técnica multivariada:.....	12
<i>III. CONCLUSIÓN</i>	17
<i>IV. REFERENCIAS</i>	18

I. INTRODUCCIÓN

En el dinámico mundo de las ventas, la capacidad de analizar y evaluar el rendimiento de los vendedores es fundamental para el éxito de cualquier organización. Este artículo se centra en un análisis multivariado exhaustivo de los vendedores. Este estudio permite una evaluación precisa y ofrece una base sólida para tomar decisiones informadas en la gestión de la fuerza de ventas.

Los datos con los que se trabajarán en este proyecto provienen de la tabla 9.12 en la página 539 del libro "Applied Multivariate Statistical Analysis, 5th Edition".

El conjunto de datos proporciona información detallada sobre diversos indicadores de rendimiento y habilidades de vendedores, capturados a través de diferentes métricas que se describen a continuación:

- x_1 = Índice de crecimiento de ventas (Index of sales growth)
- x_2 = Índice de rentabilidad de ventas (Index of sales profitability)
- x_3 = Índice de ventas a nuevas cuentas (Index of new-account sales)
- x_4 = Puntuación en la prueba de creatividad (Score on creativity test)
- x_5 = Puntuación en la prueba de razonamiento mecánico (Score on mechanical reasoning test)
- x_6 = Puntuación en la prueba de razonamiento abstracto (Score on abstract reasoning test)
- x_7 = Puntuación en la prueba de matemáticas (Score on mathematics test)

Todas las variables son numéricas y continuas. Para las variables x_1 , x_2 y x_3 se miden mediante una escala en sus índices de crecimiento, rentabilidad y ventas. Para las variables x_4 , x_5 , x_6 y x_7 se tienen las puntuaciones que obtuvieron en pruebas de razonamiento mecánico, abstracto y matemático.

Se pretende evaluar si es posible identificar grupos o clústeres dentro de los datos que agrupen a los vendedores con características y patrones similares en relación con sus indicadores de rendimiento y habilidades, además, ofrecer recomendaciones prácticas para mejorar el desempeño de los vendedores mediante el desarrollo de habilidades específicas y estrategias de venta basadas en los hallazgos obtenidos.

En la Tabla 1, se muestran las primeras 10 filas del conjunto de datos, que contiene 50 filas y 7 columnas en total.

x_1	x_2	x_3	x_4	x_5	x_6	x_7
93	96	97.8	9	12	9	20
88.8	91.8	96.8	7	10	10	15
95	100.3	99	8	12	9	26
101.3	103.8	106.8	13	14	12	29
102	107.8	103	10	15	12	32
95.8	97.5	99.3	10	14	11	21
95.5	99.5	99	9	12	9	25
110.8	122	115.3	18	20	15	51
102.8	108.3	103.8	10	17	13	31
106.8	120.5	102	14	18	11	39

Tabla 1. Primeras 10 filas del conjunto de datos.

II. DESARROLLO

A. Análisis exploratorio de los datos

El análisis exploratorio comienza con el análisis descriptivo de nuestros datos, a continuación, se pueden observar en la Tabla 2 los mínimos, cuartiles, medianas, medias y máximos del conjunto de datos.

	x1	x2	x3	x4	x5	x6	x7
Min.	81.5	87.3	94.3	1	5	5	9
1st Qu.	93.55	99.5	99.08	8.25	12	9	21.5
Median	100.65	106.2	103.15	10	15	11	31.5
Mean	98.84	106.6	102.81	11.22	14.18	10.56	29.76
3rd Qu.	105.05	114.8	106.45	14	17	12	37
Max.	110.8	122.3	115.3	18	20	15	51

Tabla 2. Estadísticas descriptivas del conjunto de datos.

De la Tabla 2 se observa que las primeras 3 variables (x1, x2 y x3) correspondientes a los índices contienen valores cercanos a cien, mientras que para las variables restantes (x4, x5, x6 y x7) los valores son en su mayoría en la escala de un par de decenas.

La matriz de correlación se utiliza para entender la variabilidad conjunta de las variables en el conjunto de datos. A continuación, se presenta en la Tabla 3 la matriz de correlación calculada para las siete variables:

	x1	x2	x3	x4	x5	x6	x7
x1	1	0.9260758	0.8840023	0.5720363	0.7080738	0.6744073	0.9273116
x2	0.9260758	1	0.8425232	0.541508	0.7459097	0.465388	0.944296
x3	0.8840023	0.8425232	1	0.700363	0.6374712	0.6410886	0.8525682
x4	0.5720363	0.541508	0.700363	1	0.590736	0.1469074	0.4126395
x5	0.7080738	0.7459097	0.6374712	0.590736	1	0.3859502	0.5745533
x6	0.6744073	0.465388	0.6410886	0.1469074	0.3859502	1	0.5663721
x7	0.9273116	0.944296	0.8525682	0.4126395	0.5745533	0.5663721	1

Tabla 3. Matriz de correlación.

La matriz de correlación revela que las variables de rendimiento de ventas (x1, x2, x3) están fuertemente correlacionadas entre sí y con la puntuación en la prueba de matemáticas (x7). Además, hay relaciones significativas entre otras variables de pruebas cognitivas y de habilidades, como la creatividad (x4), el razonamiento mecánico (x5), y el razonamiento abstracto (x6), aunque con menos fuerza en comparación con las variables de ventas. Estas relaciones pueden ser útiles para identificar qué factores contribuyen al éxito de ventas y para diseñar estrategias de capacitación y desarrollo para los vendedores.

En este análisis, se calculó el vector de medias para cada una de las siete variables en el conjunto de datos de ventas. Estos valores representan los promedios de cada variable, proporcionando una visión general de las tendencias centrales de los datos. A continuación, se presentan los resultados en la Tabla 4 y se presenta en forma vectorial:

x1	98.836
x2	106.622
x3	102.81
x4	11.22
x5	14.18
x6	10.56
x7	29.76

Tabla 4. Vectores de medias.

$$\mu = \begin{bmatrix} 98.836 \\ 106.622 \\ 102.81 \\ 11.22 \\ 14.18 \\ 10.56 \\ 29.76 \end{bmatrix}$$

La matriz de covarianza y varianza revela las relaciones significativas entre las variables en nuestro conjunto de datos, así como la variabilidad individual de cada variable. Las relaciones fuertes entre las variables de rendimiento de ventas (x1, x2, x3) y las puntuaciones en pruebas cognitivas, especialmente matemáticas (x7), sugieren que ciertos factores cognitivos pueden influir significativamente en el rendimiento de ventas. Esta información puede ser útil para diseñar programas de desarrollo y capacitación para mejorar el desempeño en ventas.

	x1	x2	x3	x4	x5	x6	x7
x1	53.83664	68.79409	30.564531	16.579673	17.585224	10.587592	71.69861
x2	68.79409	102.50175	40.195082	21.656286	25.561265	10.081306	100.74416
x3	30.56453	40.19508	22.205	13.036531	10.167551	6.463673	42.3351
x4	16.57967	21.65629	13.036531	15.603673	7.898367	1.241633	17.17633
x5	17.58522	25.56127	10.167551	7.898367	11.456735	2.795102	20.49306
x6	10.58759	10.08131	6.463673	1.241633	2.795102	4.577959	12.7698
x7	71.69861	100.74416	42.335102	17.176327	20.493061	12.769796	111.04327

Tabla 5. Matriz de varianza y covarianza.

Realizando los histogramas de cada variable nos dará una idea general de cómo se distribuyen los datos obtenidos y si tienen cierto comportamiento normal, las gráficas generadas se presentan en la Figura 1.

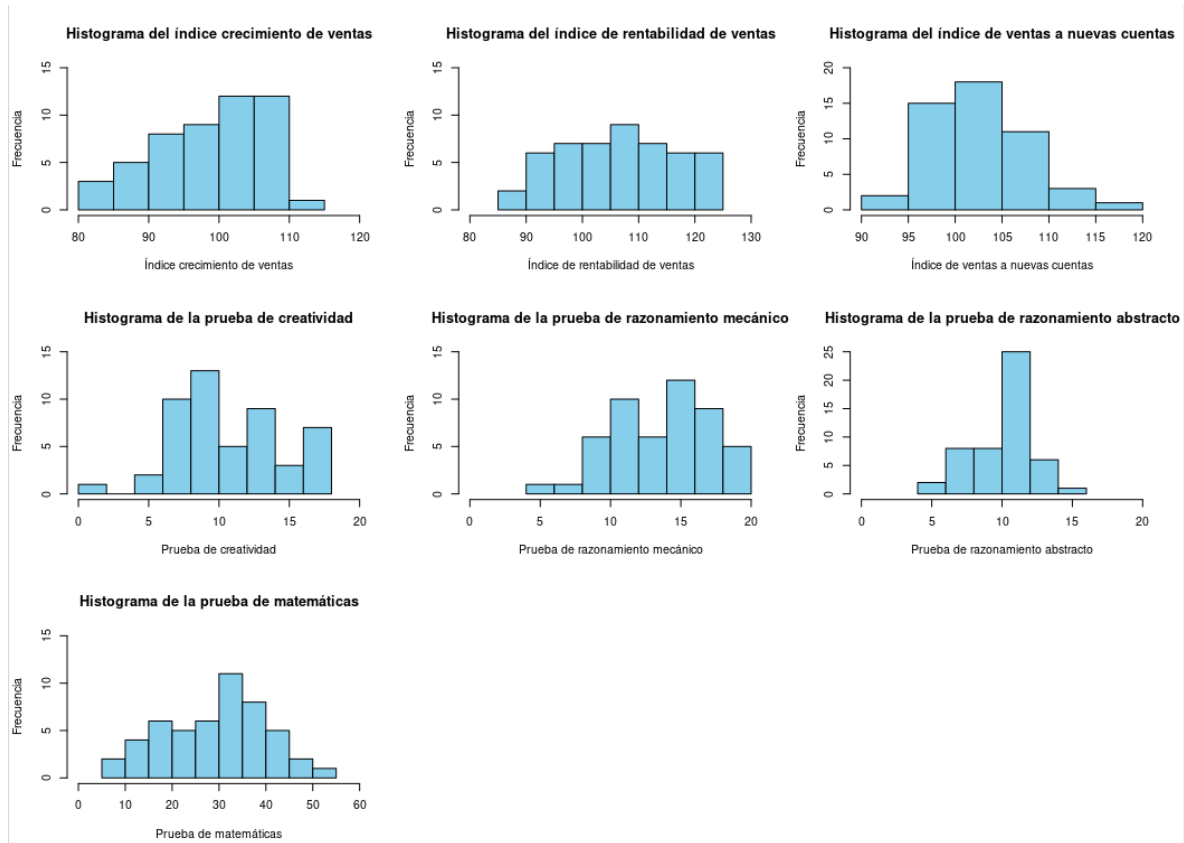


Figura 1. Histogramas de las siete variables en orden.

En este caso parecieran tener un comportamiento que no es normal, pero la forma es cercana en algunos casos a una campana de Gauss, lo cual indicaría una sospecha de normalidad para estos casos.

B. Pruebas de bondad de ajuste

Se utilizó la librería MVN en el entorno de programación R para llevar a cabo un análisis de normalidad multivariada (MVN) usando la función **mvn**. Esta librería es reconocida por su capacidad para implementar las tres pruebas MVN más prevalentes en la literatura científica: la prueba de Mardia, la prueba de Henze-Zirkler y la prueba de Royston. Estas pruebas son esenciales para verificar si un conjunto de datos multivariante sigue una distribución normal. Además de las pruebas estadísticas, MVN también proporciona algunos enfoques gráficos como chi-cuadrado Q-Q, perspectiva y gráficos de contorno. En este caso se optó por utilizar la prueba de Mardia la cual se basa en extensiones multivariadas de oblicuidad (γ_1, p) y curtosis (γ_2, p) medidas de la siguiente manera:

$$(\gamma_1, p) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n m_{ij}^3$$

$$(\gamma_2, p) = \frac{1}{n} \sum_{i=1}^n m_{ii}^2$$

Donde $m_{ij} = (x_i - \bar{x})' S^{-1} (x_i - \bar{x})$, la distancia de Mahalanobis al cuadrado, y p es el número de variables. El estadístico de prueba de oblicuidad, $\left(\frac{n}{6}\right) \gamma_1, p$ está aproximadamente χ^2 distribuido con $p(p+1)(p+2)/6$ grados de libertad. De manera similar, el estadístico de prueba para curtosis, γ_2, p , tiene una distribución aproximadamente normal con media $p(p+2)$ y varianza $8p(p+2)/n$ [2].

La Tabla 6 muestra los resultados de la prueba con una significancia de $\alpha = 0.05$

Prueba de Hipótesis	
H₀	Los datos provienen de población con distribución normal multivariada
H_a	Los datos NO provienen de población con distribución normal multivariada
Estadístico de Prueba	Oblicuidad: 138.478863289409 Curtosis: 2.53225049191325
p-valor	Oblicuidad: 0.000169934283751658 Curtosis: 0.0113332995941025
Comparativa y conclusión respecto al rechazo de H₀	Se rechaza H ₀ si p-valor < α En este caso 0.0001 < 0.05 0.01 < 0.05 Por lo tanto se rechaza H ₀
Conclusión en contexto del ejercicio	Al rechazar H ₀ , con un 95% de confianza se puede afirmar que los datos NO provienen de población con distribución normal multivariada

Tabla 6. Prueba de hipótesis para distribución normal multivariada

Dado que los resultados presentados en la Tabla 6 indican que los datos no siguen una distribución normal multivariada, se consideró necesario realizar un análisis adicional de la normalidad univariada de cada variable. Este paso es crucial para identificar si alguna de las variables individuales cumple con los supuestos de normalidad, lo cual puede tener implicaciones significativas para los métodos estadísticos a emplear posteriormente, de igual manera se utilizó la librería MVN para realizar este análisis, se empleó la prueba de Anderson-Darling con una significancia de $\alpha = 0.05$, los resultados se presentan en la Tabla 7.

Variable	Estadístico de Prueba	p-valor	Normalidad
x1	1.0433	0.0088	No
x2	0.5266	0.171	Sí
x3	0.3372	0.4916	Sí
x4	0.652	0.0838	Sí
x5	0.4681	0.2394	Sí
x6	1.3037	0.002	No
x7	0.376	0.3993	Sí

Tabla 7. Prueba de hipótesis para distribución normal multivariada

Los resultados presentados en la Tabla 7 indican que, con un nivel de confianza del 95%, la mayoría de las variables siguen una distribución normal, excepto las variables x1 y x6. Por este motivo, se decidió eliminar estas dos variables del análisis. Esta decisión se tomó con el objetivo de realizar una prueba de normalidad multivariada excluyendo estas dos variables, y así determinar si el conjunto de datos, compuesto por las cinco variables restantes, se ajusta a una distribución normal multivariada. Este enfoque nos permitirá evaluar la validez de los supuestos estadísticos necesarios para los análisis posteriores, mejorando la fiabilidad y precisión de las inferencias que se deriven de los datos.

Se realizó una vez más la prueba de Mardia una significancia de $\alpha = 0.05$, para evaluar si el conjunto de datos proviene de una distribución normal multivariada, los resultados se presentan en la Tabla 8.

Prueba de Hipótesis	
H₀	Los datos provienen de población con distribución normal multivariada
H_a	Los datos NO provienen de población con distribución normal multivariada
Estadístico de Prueba	Oblicuidad: 31.3451126088302 Curtosis: 0.0349831105952262
p-valor	Oblicuidad: 0.645257206200229 Curtosis: 0.972093208422376
Comparativa y conclusión respecto al rechazo de H₀	Se rechaza H ₀ si p-valor < α En este caso 0.64 > 0.05 0.97 > 0.05 Por lo tanto NO se rechaza H ₀
Conclusión en contexto del ejercicio	Al no rechazar H ₀ , con un 95% de confianza se puede afirmar que los datos provienen de población con distribución normal multivariada

Tabla 8. Prueba de hipótesis para distribución normal multivariada

De la Tabla 8 se observa que, con un nivel de confianza del 95%, el nuevo conjunto de datos sigue una distribución normal multivariada. Este hallazgo permite realizar análisis multivariados de manera precisa y confiable, asegurando que los supuestos estadísticos necesarios para estos métodos se cumplen adecuadamente.

C. Planteamiento de comparativas:

Para determinar si las variables son independientes entre sí se realizó una prueba de varianzas de Bartlett, donde

$$\Sigma_0 = \begin{bmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{bmatrix}$$

Ya que para que para que 2 variables sean independientes entre sí, su covarianza debe ser igual a cero. En la Tabla 9 se presentan los resultados de la prueba de hipótesis con una significancia de $\alpha = 0.05$

Prueba de Hipótesis	
H₀	Las 5 variables son independientes entre sí
H_a	Al menos un par de variables no son independientes entre sí
Estadístico de Prueba (Bartlett)	287.1494
Región de rechazo	24.99579
Comparativa y conclusión respecto al rechazo de H₀	Se rechaza H ₀ si el estadístico de prueba > Región de rechazo En este caso 287.1494 > 24.99579 Por lo tanto, se rechaza H ₀
Conclusión en contexto del ejercicio	Al rechazar H ₀ , con un 95% de confianza se puede afirmar que al menos un par de variables no son independientes entre sí

Tabla 9. Prueba de hipótesis para la independencia entre las variables

Según los resultados presentados en la tabla 9, con un nivel de confianza del 95%, se observa que existe evidencia estadística que sugiere que al menos dos variables no son independientes entre sí.

La empresa a la cual los vendedores pertenecen indica que el promedio para los índices es de 100 puntos, mientras que para las pruebas son 15 puntos y en el caso de la prueba de matemáticas son 30 puntos, por lo tanto, se busca conocer si existe evidencia estadística para afirmar si el vector de medias correspondientes a las variables medidas puede ser considerado como 30; respectivamente.

Se realizó una prueba de hipótesis para la comparación de un vector de medias con una significancia de $\alpha = 0.05$, los resultados se presentan en la Tabla 10.

Prueba de Hipótesis	
H₀	El vector de medias está dado por $\mu'_0 = [100 \ 100 \ 15 \ 15 \ 30]$
H_a	El vector de medias difiere de: $\mu'_0 = [100 \ 100 \ 15 \ 15 \ 30]$
Estadístico de Prueba (T² de Hotelling)	1804.227
Región de rechazo	2.422085
Comparativa y conclusión respecto al rechazo de H₀	Se rechaza H ₀ si el estadístico de prueba > Región de rechazo En este caso $1804.227 > 2.422085$ Por lo tanto, se rechaza H ₀
Conclusión en contexto del ejercicio	Al rechazar H ₀ , con un 95% de confianza se puede afirmar que El vector de medias difiere de: $\mu'_0 = [100 \ 100 \ 15 \ 15 \ 30]$

Tabla 10. Prueba de hipótesis para un vector de medias

La Tabla 10 presenta los resultados de la prueba de hipótesis realizada para evaluar si el vector de medias del conjunto de datos coincide con un vector de referencia predefinido. Los resultados muestran que se rechazó la hipótesis nula, lo que indica una diferencia significativa entre el vector de medias observado y el vector de referencia. Este hallazgo sugiere que al menos una de las medias de las variables no coincide con la referencia

esperada. No obstante, una limitación importante de esta prueba es que no permite determinar específicamente cuáles valores del vector de medias no cumplen con el supuesto, ya que el análisis se realiza sobre el vector completo.

Además, se busca determinar si se puede considerar si los promedios de calificación por prueba no difieren significativamente entre sí, solo para los resultados de las pruebas, por lo que en este análisis solo se consideran las variables x_4 , x_5 , y x_7 . Se realizó una prueba MANOVA con una significancia de $\alpha = 0.05$ para determinar si existe diferencia entre la media de los puntajes obtenidos en las distintas pruebas. Los resultados se presentan en la Tabla 11.

Prueba de Hipótesis	
H_0	Las medias de puntaje por prueba no difieren
H_a	Al menos una media difiere
Estadístico de Prueba	86.00494
Región de rechazo	2.655939
Comparativa y conclusión respecto al rechazo de H_0	Se rechaza H_0 si el estadístico de prueba > Región de rechazo En este caso $86.00494 > 2.655939$ Por lo tanto, se rechaza H_0
Conclusión en contexto del ejercicio	Al rechazar H_0 , con un 95% de confianza se puede afirmar que al menos una prueba difiere.

Tabla 11. Prueba de hipótesis para un vector de medias

Como se observa en la Tabla 11, con un nivel de confianza del 95%, se puede afirmar que al menos uno de los promedios difiere significativamente de los demás. Esto sugiere que existe una variación estadísticamente significativa entre los promedios de las diferentes pruebas, ya que las pruebas están diseñadas para evaluar amplias aptitudes en lugar de enfocarse en una sola área específica.

D. Técnica multivariada:

Se llevó a cabo un análisis de componentes principales (PCA) con el objetivo de evaluar la posibilidad de reducir la dimensionalidad del conjunto de datos. Este enfoque estadístico se utilizó para identificar los componentes subyacentes que capturan la mayor parte de la variabilidad presente en los datos originales. Al transformar las variables originales en un nuevo conjunto de variables no correlacionadas (componentes principales), se buscó simplificar la estructura del conjunto de datos, preservando la mayor cantidad de información posible. Esta reducción de dimensionalidad no solo facilita la interpretación y visualización de los datos, sino que también puede mejorar la eficiencia de los modelos predictivos y las estrategias de análisis posteriores.

Se empleó una gráfica de codo presentada en la Figura 2 para determinar la cantidad óptima de componentes principales (PCA) a retener. Esta técnica visual permite identificar el punto en el cual la adición de componentes adicionales no contribuye significativamente a la explicación de la varianza total de los datos. Al analizar la gráfica, se puede observar el punto de inflexión que indica la cantidad óptima de componentes a conservar, equilibrando la complejidad del modelo y la cantidad de información retenida

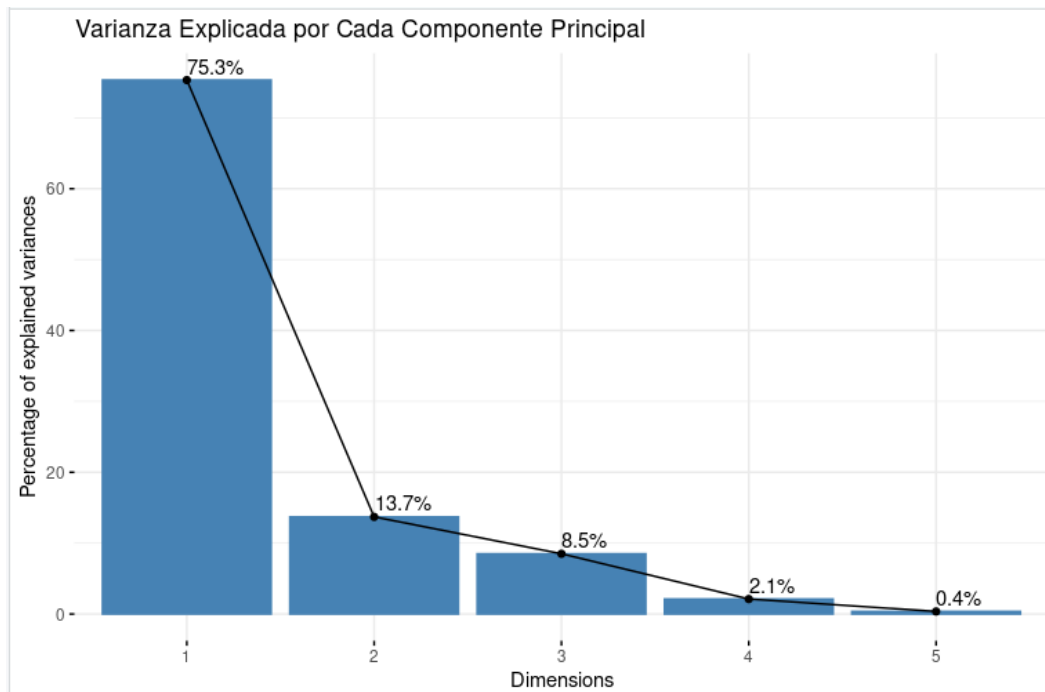


Figura 2. Grafica de codo.

Componentes Principales: En la Figura 2 se muestra que PC1 explica el 75.33% de la varianza, mientras que PC2 añade un 13.70%.

Varianza: De la Figura 2 se observa que los primeros dos componentes principales (PC1 y PC2) explican el 89.03% de la varianza total en los datos. Esto indica que la mayor parte de la variabilidad en el conjunto de datos puede ser capturada por solo dos componentes.

Según los resultados obtenidos, es posible reducir las dimensiones del conjunto de datos de 5 a 2 componentes principales sin perder una cantidad significativa de información, ya que estos dos componentes capturan el 89.03% de la varianza total.

Después de utilizar únicamente dos componentes principales los coeficientes asociados a la transformación se presentan en la Tabla 12

	PC1	PC2
x2	0.490002	0.2922634
x3	0.4822992	0.02361078
x4	0.3758768	-0.76435096
x5	0.4188066	-0.24977251
x7	0.4589152	0.51711308

Tabla 12. Coeficientes asociados a la transformación.

Una vez obtenidos los coeficientes asociados a la transformación mediante el análisis de componentes principales (PCA), se procedió a realizar la transformación del conjunto de datos. Este proceso implicó la multiplicación matricial de los datos originales con los coeficientes de la transformación de la Tabla 12. De esta manera, se generó un nuevo conjunto de datos de únicamente 2 variables, donde las dimensiones originales se proyectan en un menor número de componentes no correlacionados que capturan la mayor parte de la variabilidad. Este nuevo conjunto de datos facilita una interpretación más clara y la aplicación de técnicas analíticas subsecuentes con mayor eficiencia, en la Tabla 13 se presentan las primeras 10 filas del conjunto de datos transformado:

	PC1	PC2
1	-1.93289361	-0.22024527
2	-2.89404215	-0.05727813
3	-1.43581518	0.39783265
4	0.38580536	-0.42991319
5	0.15938298	0.32043961
6	-1.32060011	-0.46144058
7	-1.42292872	0.13216674
8	4.31290566	-0.19259098
9	0.46937767	0.14222347
10	1.72835964	0.03017877

Tabla 13. Datos transformados.

Este análisis sugiere que se puede simplificar el conjunto de datos original mientras se mantiene la mayor parte de la información, facilitando análisis posteriores y visualización de datos. Ahora que la dimensionalidad del conjunto de datos se ha reducido a dos dimensiones, se procedió a graficar los dos componentes principales como se muestra en la Figura 3. Esta representación visual permite observar cómo las observaciones se distribuyen en el espacio definido por los componentes principales, facilitando la identificación de patrones, agrupamientos y posibles outliers.

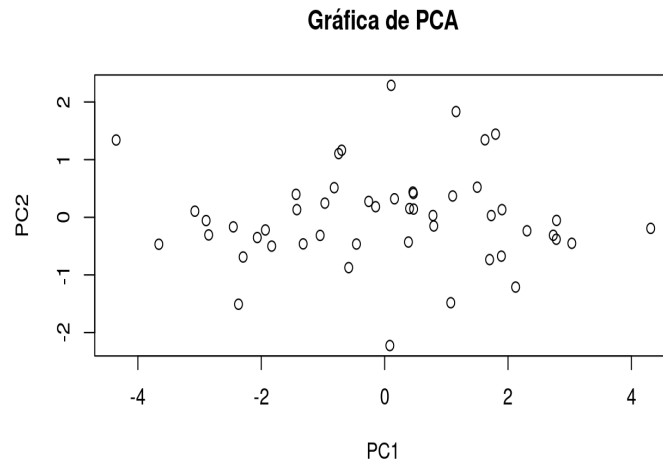


Figura 3. Grafica de PCA.

Análisis de conglomerado

Se eligió el análisis de conglomerados como la primera técnica multivariada debido a su capacidad para identificar y agrupar observaciones con características similares. Esta técnica ofrece una visión clara de cómo se agrupan las diferentes observaciones, lo cual es crucial para una comprensión más profunda y detallada del comportamiento y las competencias de los vendedores.

1. **Calcular la Distancia Euclidiana:** Se utilizó la distancia euclidiana para medir la similitud entre las observaciones del conjunto de datos escalados.
2. **Clustering Jerárquico:** Se aplicó el método de Ward para realizar el clustering jerárquico, el cual minimiza la varianza dentro de los clusters.
3. **Dendrograma:** Se generó un dendrograma para visualizar el agrupamiento jerárquico de las observaciones. Esto ayuda a decidir el número óptimo de clusters.
4. **Definir Clusters:** Se cortó el dendrograma en cuatro grupos. Este número fue seleccionado basado en la estructura observada en el dendrograma.

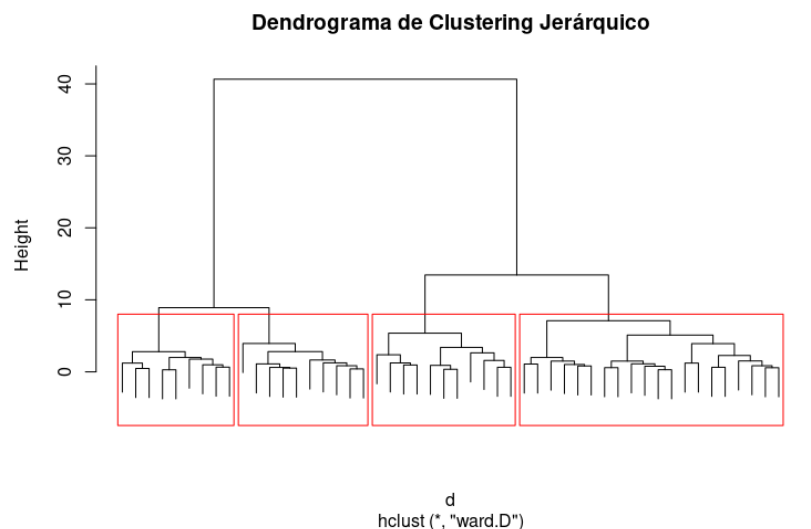


Figura 4. Dendrograma de los grupos creados.

5. **Asignar Clusters al DataFrame:** Los grupos resultantes se añadieron al DataFrame original, permitiendo una mejor interpretación y análisis posterior.

La Figura 4 presenta el dendrograma resultante del análisis de conglomerados, revelando la estructura jerárquica de los grupos formados. En este dendrograma se pueden identificar claramente cuatro grupos principales, lo cual proporciona una visualización intuitiva de cómo se relacionan las diferentes observaciones y cómo se agrupan en función de sus similitudes.

Resultados:

El análisis de conglomerados resultó en la formación de cuatro grupos distintos, cada uno representando un conjunto de observaciones con características similares. A continuación, se presenta un ejemplo de las observaciones agrupadas:

- **Grupo 1:** Caracterizado por observaciones con valores más bajos en la mayoría de las variables.
- **Grupo 2:** Observaciones con valores moderados y consistentes en las variables.
- **Grupo 3:** Incluye observaciones con valores más altos en algunas variables específicas.
- **Grupo 4:** Representa el grupo con los valores más altos en la mayoría de las variables, lo cual puede indicar un perfil particular en los datos.

Estos grupos permiten una mejor comprensión de la estructura subyacente en los datos, facilitando la identificación de patrones y la toma de decisiones basada en el comportamiento observado en cada conglomerado [3].

Con estos resultados podemos tomar decisiones para el mejoramiento del desempeño de los vendedores que no pertenecen al grupo 4, por ejemplo, incentivos económicos si pertenecen al grupo o apoyo para tomar certificaciones o cursos que los ayuden a mejorar.

En la Tabla 14 se presentan los primeros 10 registros con el grupo asignado mediante el análisis de conglomerados.

	x2	x3	x4	x5	x7	Grupo
1	96	97.8	9	12	20	1
2	91.8	96.8	7	10	15	1
3	100.3	99	8	12	26	2
4	103.8	106.8	13	14	29	3
5	107.8	103	10	15	32	3
6	97.5	99.3	10	14	21	2
7	99.5	99	9	12	25	2
8	122	115.3	18	20	51	4
9	108.3	103.8	10	17	31	3
10	120.5	102	14	18	39	3

Tabla 14. Datos con su grupo asignado.

Análisis factorial

Se eligió el análisis de factores como segunda técnica multivariada debido a su capacidad para identificar las dimensiones subyacentes que explican las correlaciones entre las variables observadas. Esta técnica es útil para reducir la cantidad de variables en un

conjunto de datos complejo y para descubrir factores latentes que influyen en el comportamiento de las observaciones. Al aplicar el análisis de factores, se pueden identificar las variables más importantes y cómo se agrupan en factores, lo que facilita una comprensión más profunda de las relaciones entre las diferentes métricas y habilidades de los vendedores. Esta técnica complementa el análisis de conglomerados, proporcionando una perspectiva adicional sobre las estructuras subyacentes en los datos y ayudando a diseñar estrategias de intervención más efectivas basadas en factores clave.

Selección del Número de Factores: Se decidió utilizar 2 factores para el análisis, basándonos en la gráfica de codo.

Análisis de Componentes Principales: Se aplicó la rotación Varimax para simplificar la interpretación de los factores.

Matriz de Cargas Estandarizadas: La matriz de cargas muestra cómo cada variable original se relaciona con los factores obtenidos:

- RC1: Alta carga en las variables x2, x3 y x7.
- RC2: Alta carga en las variables x3, x4 y x5.

Proporción de Varianza Explicada:

- Factor 1 (RC1): Explica el 53.1% de la varianza total.
- Factor 2 (RC2): Explica el 35.9% de la varianza total.
- Cumulativo: Los dos factores juntos explican el 89% de la varianza total.

Comunidades: Indican la proporción de varianza de cada variable explicada por los factores.

x2	x3	x4	x5	x7
0.9628914	0.8765369	0.9324619	0.7034013	0.9764769

Especificidades: Varianza específica no explicada por los factores.

x2	x3	x4	x5	x7
0.03710864	0.12346312	0.06753809	0.29659869	0.02352314

Resumen de Resultados:

- **Cargas significativas:** Variables x2 y x7 cargan fuertemente en RC1, mientras que x4 y x5 lo hacen en RC2.
- **Proporción de varianza:** RC1 y RC2 explican conjuntamente el 89% de la varianza total del conjunto de datos.

Interpretación de los Factores:

- **Factor 1 (RC1):** Representa una combinación de variables con alta carga en x2, x3 y x7, sugiriendo un factor subyacente relacionado con estas dimensiones.
- **Factor 2 (RC2):** Representa una combinación de variables con alta carga en x3, x4 y x5, sugiriendo otro factor subyacente relacionado con estas dimensiones.

III. CONCLUSIÓN

El análisis del conjunto de datos ha proporcionado una visión integral sobre las características y el desempeño de los vendedores en relación con diversas métricas y habilidades. El análisis realizado proporciona una visión valiosa sobre cómo las habilidades específicas y los indicadores de rendimiento en ventas están interrelacionados. Los hallazgos sugieren que las habilidades matemáticas tienen un impacto notable en el rendimiento en ventas, mientras que otras habilidades cognitivas tienen un efecto más limitado. Los grupos identificados y los factores subyacentes ofrecen una base para diseñar estrategias de capacitación más dirigidas y mejorar el rendimiento general de los vendedores mediante el desarrollo de habilidades clave.

Se observó que las pruebas, aunque diseñadas para evaluar una amplia gama de aptitudes, presentan variaciones en su capacidad para medir las habilidades de los vendedores. Esta heterogeneidad en los resultados es coherente con el objetivo diversificado de las pruebas, que abordan diferentes aspectos del desempeño y las competencias, lo cual contribuye a la variabilidad en los promedios obtenidos.

La evaluación de la posibilidad de agrupamiento reveló que es factible identificar clusters dentro de los datos, lo cual permite agrupar a los vendedores en función de características y patrones similares. Esta segmentación proporciona una base para personalizar estrategias de desarrollo y optimizar el rendimiento en ventas.

Además, las recomendaciones basadas en los hallazgos sugieren que invertir en el desarrollo de habilidades específicas, como el razonamiento mecánico y abstracto, podría ser beneficioso para mejorar el desempeño general. Además, las estrategias personalizadas para los diferentes clusters identificados pueden ayudar a maximizar la efectividad de los vendedores y potenciar el éxito de ventas. Por lo tanto, el análisis realizado no solo ha demostrado la existencia de relaciones significativas entre las variables, sino que también ha proporcionado una comprensión más profunda de cómo diferentes habilidades impactan el desempeño en ventas, ofreciendo así una guía valiosa para la toma de decisiones y el diseño de estrategias efectivas.

En conclusión, el estudio ha subrayado la importancia de adoptar un enfoque multivariado para analizar el rendimiento de los vendedores, destacando su papel crucial en la optimización de la efectividad de la fuerza de ventas. Las metodologías empleadas en el análisis proporcionan una evaluación detallada y precisa de las múltiples dimensiones del desempeño de los vendedores. Además, estas técnicas no solo facilitan una comprensión profunda de los factores que influyen en el éxito de ventas, sino que también ofrecen estrategias prácticas y basadas en datos para fomentar la mejora continua del desempeño en diversos contextos empresariales. Este enfoque integral permite a las organizaciones desarrollar intervenciones más específicas y efectivas, ajustadas a las necesidades particulares de sus equipos de ventas y adaptadas a las dinámicas cambiantes del mercado.

IV. REFERENCIAS

- [1] Johnson, R. A., & Wichern, D. W. (2002). Applied multivariate statistical analysis.
- [2] Korkmaz, Selcuk, et al. "MVN: An R Package for Assessing Multivariate Normality." The R Journal, vol. 6, no. 2, 2014, p. 151, <https://doi.org/10.32614/rj-2014-031>.
- [3] Aranibar, G. R. (2019). Análisis de conglomerados. Revista Varianza, 65-84.