

# Análisis de Factores Fisiológicos, Demográficos y Ambientales para Evaluar el Rendimiento en Ejercicio Máximo

José A. López-Torres<sup>1</sup>, Diana L. Hernández-Almeida<sup>1</sup> and Axel D. Luna-Hernández<sup>1</sup>

<sup>1</sup> Tecnológico de Monterrey, Campus Guadalajara, Ingeniería en Ciencia de Datos

**Abstract**— El rendimiento durante la prueba de Ejercicio Máximo fue evaluado considerando factores fisiológicos, demográficos y ambientales. Se analizaron variables fisiológicas como la frecuencia cardíaca, el consumo de oxígeno, la producción de dióxido de carbono y la ventilación pulmonar; condiciones técnicas de la prueba, como la velocidad de la cinta; además de características individuales, como la edad, el peso y el género, junto con condiciones ambientales, como la temperatura y la humedad. El análisis de componentes principales (PCA) fue aplicado para reducir la dimensionalidad de los datos, lo que facilitó la identificación de patrones significativos en las respuestas de los participantes. Posteriormente, fueron implementados métodos de agrupamiento para clasificar a los individuos según sus características y comportamiento durante el esfuerzo físico. Resultados indicaron que las variables fisiológicas tuvieron una contribución significativa al rendimiento, mientras que las condiciones demográficas y ambientales no influenciaron los patrones de respuesta de manera uniforme, en específico la edad no hace gran diferencia y las condiciones ambientales influyen de manera moderada. De esta manera, grupos de participantes con perfiles específicos fueron identificados, permitiendo que estrategias de entrenamiento sean personalizadas y que la comprensión de las respuestas al ejercicio bajo diferentes condiciones sea mejorada. Estos hallazgos proporcionan una base para que programas de entrenamiento más efectivos sean desarrollados, adaptados a las características únicas de cada individuo, y para que el rendimiento físico en diversas poblaciones y entornos sea optimizado.

**Keywords**— Rendimiento físico, Análisis de componentes principales (PCA), Agrupamiento, Factores fisiológicos, Factores demográficos, Factores ambientales, Ejercicio, Optimización de entrenamiento

## I. INTRODUCCIÓN

El rendimiento físico durante pruebas de ejercicio ha sido ampliamente investigado debido a su importancia en áreas como el deporte, la salud y la rehabilitación. Se ha determinado que comprender cómo diferentes factores influyen en el desempeño físico es esencial para que entrenamientos sean optimizados, lesiones sean prevenidas y la calidad de vida en la población sea mejorada. Las variables fisiológicas, como la frecuencia cardíaca, el consumo de oxígeno y la ventilación pulmonar, han sido destacadas como indicadores clave del desempeño físico. Sin embargo, factores demográficos, como la edad y el género, al igual que condiciones ambientales, como la temperatura y la humedad, también han sido identificados como elementos con influencia significativa en las respuestas al esfuerzo físico.

Es importante señalar que estas tres áreas suelen ser analizadas por separado. Por ejemplo, investigaciones recientes han señalado que la variabilidad respiratoria durante el ejercicio es influenciada por la intensidad del esfuerzo y las diferencias entre géneros, observándose que la variabilidad en la ventilación pulmonar (VE) es menor en mujeres en comparación con los hombres, independientemente del nivel de intensidad del ejercicio o del uso de reservas respiratorias [1]. A pesar de estos avances, el análisis conjunto de estas variables

continúa representando un desafío debido a la complejidad dimensional de las mismas. En la mayoría de los estudios previos, estas variables han sido abordadas de manera aislada o han incluido un número limitado de factores, lo que ha limitado la comprensión de su impacto combinado en el rendimiento físico. Adicionalmente, enfoques individuales no han sido efectivos para identificar patrones en conjuntos de datos relacionados con pacientes o poblaciones en general, lo que dificulta que estrategias personalizadas sean diseñadas en base a características individuales. Estas limitaciones resaltan la necesidad de que métodos analíticos avanzados sean aplicados para reducir la dimensionalidad, clasificar a los participantes en grupos significativos y modelar las correlaciones entre las variables clave.

En este contexto, el presente estudio fue diseñado para analizar cómo factores fisiológicos, demográficos y ambientales influyen en el rendimiento durante pruebas de ejercicio máximo. Para ello, el análisis de componentes principales (PCA) fue empleado para simplificar los datos, métodos de agrupamiento (*clustering*) fueron implementados para que patrones significativos en las respuestas de los participantes fueran identificados, y análisis de regresión principal (PCR) fue realizado para modelar la relación entre las variables clave y el rendimiento. Este enfoque busca que vacíos en los análisis existentes sean llenados, proporcionando her-

ramientas que permitan que planes de entrenamiento sean personalizados y que el rendimiento físico en diversas poblaciones y entornos sea optimizado. Se espera que los resultados contribuyan a que programas en el área de bienestar sean diseñados de manera más efectiva, adaptados a las características únicas de cada individuo, y que la comprensión de las respuestas al ejercicio bajo diferentes condiciones ambientales sea mejorada.

## II. METODOLOGÍA

El análisis fue realizado utilizando un conjunto de datos proporcionado por PhysioNet [2],[3], el cual incluyó variables demográficas como edad, peso, estatura y género; variables ambientales como humedad y temperatura; y variables fisiológicas relacionadas con el ejercicio, como frecuencia cardíaca (HR), consumo de oxígeno (VO<sub>2</sub>), producción de dióxido de carbono (VCO<sub>2</sub>), frecuencia respiratoria (RR) y ventilación pulmonar (VE).

Una descripción detallada de estas variables se presenta en la Tabla 1. Los procedimientos principales incluyeron la imputación de datos faltantes, la estandarización de las variables, la reducción de dimensionalidad mediante el análisis de componentes principales (PCA) y la agrupación de patrones significativos por técnicas de *clustering*.

Posteriormente, se llevó a cabo un análisis de correlaciones para interpretar las relaciones entre variables clave y fue desarrollado un modelo de regresión principal (PCR) con el propósito de evaluar el impacto de estas variables en el rendimiento físico.

### a. Imputación de Datos

Los conjuntos de datos iniciales (subject-info.csv y test\_measure.csv) fueron identificados como incompletos, ya que contenían valores faltantes en varias variables. Para que esta limitación fuera abordada en el PCA, fue utilizada la imputación con *KNNImputer*, empleando  $k = 5$  como número de vecinos. Este método fue aplicado para estimar los valores faltantes basándose en los datos más próximos, manteniendo las relaciones intrínsecas entre las variables. Una vez que los datos fueron imputados, se verificó que no existieran valores ausentes en las bases procesadas.

### b. Exploración Inicial

La exploración inicial consistió en el análisis de las medias y varianzas de las variables en ambos conjuntos de datos. Los resultados fueron presentados en las Tablas 2, 3, 4 y 5. Este análisis preliminar permitió que variaciones importantes en las escalas de las variables fueran identificadas, lo que justificó que la estandarización fuera realizada antes de aplicar métodos como el análisis de componentes principales (PCA).

TABLE 2: MEDIA DE CADA VARIABLE EN SUBJECT INFO

Variable	Media
Age	28.979133
Weight	73.383367
Height	174.913508
Humidity	48.206835
Temperature	22.819817
Sex	0.150202
ID	424.890121

TABLE 3: VARIANZA DE CADA VARIABLE EN SUBJECT INFO

Variable	Varianza
Age	101.538928
Weight	144.128694
Height	63.202925
Humidity	71.564071
Temperature	7.553830
Sex	0.127770
ID	59454.277521

TABLE 4: MEDIA DE CADA VARIABLE EN TEST MEASURE

Variable	Media
time	628.126172
Speed	9.607958
HR	146.859737
VO <sub>2</sub>	2311.121681
VCO <sub>2</sub>	2387.100996
RR	36.491513
VE	69.304931
ID	425.424579

TABLE 5: VARIANZA DE CADA VARIABLE EN TEST MEASURE

Variable	Varianza
time	1.060081e+05
Speed	2.043387e+01
HR	1.040495e+03
VO <sub>2</sub>	9.587907e+05
VCO <sub>2</sub>	1.346684e+06
RR	1.482853e+02
VE	1.126178e+03
ID	6.024251e+04

Para garantizar una escala uniforme entre las variables, se aplicó el método de estandarización utilizando *StandardScaler*. Este proceso ajustó los datos a una distribución con media igual a cero y desviación estándar igual a uno, utilizando la fórmula:

$$z = \frac{x - \mu}{\sigma}$$

donde:

- $z$ : Valor estandarizado.
- $x$ : Valor original de la variable.
- $\mu$ : Media de la variable.

**TABLE 1:** DESCRIPCIÓN DE LAS VARIABLES UTILIZADAS EN EL ANÁLISIS

Categoría	Variable	Unidad	Descripción
Fisiológicas	Frecuencia cardíaca	Latidos por minuto	Ritmo cardíaco durante el ejercicio.
	VO2	ml/kg/min	Consumo de oxígeno.
	Ventilación pulmonar	L/min	Cantidad de aire movido por los pulmones.
Demográficas	Edad	Años	Edad del participante.
	Peso	kg	Peso corporal.
	Altura	cm	Altura del participante.
Ambientales	Temperatura	°C	Temperatura ambiente durante la prueba.
	Humedad	%	Humedad relativa.

- $\sigma$ : Desviación estándar de la variable.

Este procedimiento fue diseñado para que el sesgo de magnitudes, que podría sobreponerse a otras, fuera eliminado, asegurando que todas las variables tuvieran la misma importancia en los análisis posteriores.

En los datos demográficos (`subject-info.csv`) y en las mediciones (`test_measure.csv`) se excluyeron las variables no numéricas ID, ID\_test y Sex del proceso de estandarización, ya que no aportan significado en términos de escala.

### c. Análisis de Componentes Principales PCA por separado

El PCA fue utilizado para que la dimensionalidad de los datos fuera reducida, preservando la mayor proporción de varianza posible. Este método fue aplicado para transformar las variables originales en un conjunto de componentes principales que condensaron las características esenciales del conjunto de datos.

El modelo PCA fue ajustado de forma independiente para cada base de datos, permitiendo que ambas fuentes de información fueran analizadas por separado y que las características principales de cada conjunto fueran identificadas. Para una buena visualización de la reducción, fueron generados mapas de calor, los cuales reflejaron las contribuciones de las variables a cada componente principal. Asimismo, fueron calculados y analizados tanto el porcentaje de varianza explicada por cada componente como el porcentaje de varianza acumulada. Estas visualizaciones fueron presentadas en gráficas de barras que evidenciaron qué proporción de la varianza total fue capturada por los primeros componentes y ayudaron a determinar cuántos de ellos eran suficientes para un análisis robusto y confiable.

Una vez que el modelo fue entrenado, los datos originales fueron transformados para que su dimensionalidad fuera reducida, asegurando que la mayor cantidad posible de información relevante fuera conservada. Para que la precisión del procedimiento fuera verificada, las proyecciones resultantes fueron reconstruidas, confirmando que la estructura esencial de los datos fue preservada, tal como fue mostrado en las Tablas 6, 7, 8 y 9.

La proporción de varianza explicada por cada componente se calculó utilizando la ecuación:

$$\text{Varianza Explicada} = \frac{\lambda_i}{\sum_{j=1}^p \lambda_j}$$

donde:

- $\lambda_i$  representa el valor propio del componente  $i$ .
- $\sum_{j=1}^p \lambda_j$  es la suma de los valores propios de todos los componentes.

Este análisis se llevó a cabo de manera independiente para cada base de datos.

La aplicación del PCA a cada archivo permitió identificar los componentes principales que capturan la mayor parte de la variabilidad significativa en el conjunto de datos. Las cargas de los componentes principales fueron examinadas para interpretar qué variables contribuyeron más significativamente a cada componente y poder optimizar los procesos.

**TABLE 6:** VALORES ORIGINALES (TEST\_MEASURE)

time	Speed	HR	VO2	VCO2	RR	VE
0.0	5.0	63.0	478.0	360.0	27.0	13.3
2.0	5.0	75.0	401.0	295.0	22.0	10.3
4.0	5.0	69.0	390.0	319.0	20.0	12.2
8.0	5.0	87.0	417.0	344.0	28.0	14.6

**TABLE 7:** VALORES RECONSTRUIDOS (TEST\_MEASURE)

time	Speed	HR	VO2	VCO2	RR	VE
-5.6843e-13	5.0	63.0	478.0	360.0	27.0	13.3
2.0000e+00	5.0	75.0	401.0	295.0	22.0	10.3
4.0000e+00	5.0	69.0	390.0	319.0	20.0	12.2
7.0000e+00	5.0	87.0	417.0	344.0	28.0	14.6

**TABLE 8:** VALORES ORIGINALES (SUBJECT\_INFO)

Age	Weight	Height	Humidity	Temperature
10.8	48.8	163.0	39.0	20.7
11.8	41.0	150.0	41.0	22.3
12.2	46.0	160.0	37.0	21.5
13.2	71.0	190.0	40.0	25.3

**TABLE 9:** VALORES RECONSTRUIDOS (SUBJECT\_INFO)

Age	Weight	Height	Humidity	Temperature	Sex
10.8	48.8	163.0	39.0	20.7	1
11.8	41.0	150.0	41.0	22.3	1
12.2	46.0	160.0	37.0	21.5	0
13.7	53.8	190.7	40.0	25.3	0

#### d. Análisis de Componentes Principales PCA en conjunto

Se ejecutó un tercer Análisis de Componentes Principales (PCA) después de que las variables fueran unidas en una nueva colección de datos. Para ello, las bases de datos fueron combinadas mediante la identificación de líneas similares, utilizando las variables de ID\_test como criterio de similitud. Se filtraron los *dataframes* para mantener únicamente los IDs comunes, abordando así las discrepancias entre las bases.

Una vez que las bases de datos fueron unidas, se seleccionaron las variables numéricas, y estas fueron estandarizadas, tal y como fue aplicado en los conjuntos separados. Posteriormente, se determinó el número de componentes principales necesarios para explicar el 95% de la varianza, resultando en un total de 11 componentes. Finalmente, el PCA fue aplicado utilizando el número óptimo de componentes identificados.

#### e. Métodos de Agrupamiento Clustering

El agrupamiento fue realizado mediante *K-Means* sobre las proyecciones obtenidas del PCA. Se evaluaron métricas como el método del codo para determinar el número óptimo de *clusters* y el coeficiente de silueta para diferentes cantidades de *clusters*. Este análisis permitió que patrones entre los participantes fueran identificados y que estos fueran clasificados en grupos con características diferenciadas.

Las correlaciones entre las variables seleccionadas y los *clusters* identificados fueron calculadas. Lo anterior fue visualizado mediante un gráfico de dispersión y un *heatmap*, lo que permitió que las relaciones más significativas fueran resaltadas y que el impacto de cada variable en los patrones observados fuera comprendido. Para que esta información pudiera ser utilizada en una regresión principal, fue formada una matriz de correlaciones, que sirvió como el nuevo *dataframe*.

Para finalizar la sección de análisis de *clusters*, se desarrolló una función destinada a calcular tanto la correlación como el *p-valor* entre las variables seleccionadas y los *clusters* identificados. Posteriormente, las correlaciones fueron filtradas considerando únicamente aquellas con un *p-valor* menor a 0.05, garantizando su significancia estadística. Las correlaciones resultantes fueron ordenadas en función del valor absoluto de su coeficiente, permitiendo que las correlaciones más fuertes fueran destacadas y analizadas en mayor detalle, como se evidenció en la Tabla 10.

**TABLE 10:** CORRELACIONES ENTRE VARIABLES SELECCIONADAS

Variable 1	Variable 2	Correlación	P-valor	Correlación_abs
PC5	Weight	0.924399	0.0	0.924399
Weight	PC5	0.924399	0.0	0.924399
VO2	PC1	0.914698	0.0	0.914698
PC1	VO2	0.914698	0.0	0.914698
VE	PC1	0.900826	0.0	0.900826
PC1	VE	0.900826	0.0	0.900826
PC1	VCO2	0.895415	0.0	0.895415
VCO2	PC1	0.895415	0.0	0.895415
PC5	Height	0.893492	0.0	0.893492
Height	PC5	0.893492	0.0	0.893492

Se creó un diccionario para almacenar los resultados del análisis de varianza (ANOVA). Posteriormente, las variables fueron filtradas considerando únicamente aquellas que presentaban diferencias significativas entre los *clusters*, con un *p-valor* menor a 0.05. Estas variables con diferencias significativas entre *clusters* fueron identificadas y almacenadas para análisis posteriores, como se demostró en la Tabla 11.

**TABLE 11:** VARIABLES CON DIFERENCIAS SIGNIFICATIVAS ENTRE CLUSTERS

Variable	F-Statistic	P-valor
Cluster	inf	0.000000e+00
PC1	814.723655	7.978154e-210
VO2	754.974431	8.581186e-200
VCO2	620.021912	3.091636e-175
VE	562.185685	3.279455e-156
PC2	302.856862	2.497261e-110
Sex	301.867202	2.620299e-109
PC5	285.181305	1.603111e-98
Height	236.868731	9.087258e-85
Speed	228.243532	1.344808e-82
Age	218.915632	3.197972e-79
Weight	195.137710	1.693162e-72
HR	162.907489	7.614585e-62
time	131.900663	3.084794e-52
PC6	119.406601	4.883802e-47
PC7	97.484559	4.804590e-38
RR	59.280637	8.433882e-25
Temperature	40.091451	3.143856e-16
Humidity	19.115223	1.915292e-04
PC3	7.489365	5.912284e-04
PC9	4.110779	1.667596e-02

#### f. Regresión Principal PCR

Un modelo de Regresión Principal (PCR) fue desarrollado con el objetivo de predecir el consumo de oxígeno (VO2) en función de las variables previamente seleccionadas por el anterior modelo de correlaciones.

La relación entre las variables dependiente e independientes fue modelada utilizando la ecuación general de regresión principal:

$$\hat{y} = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \dots + \beta_m z_m$$

donde:

- $\hat{y}$  representa la variable dependiente (VO2).
- $\beta_0$  es el término independiente del modelo de regresión.
- $z_i$  denota el  $i$ -ésimo componente principal utilizado como predictor.
- $\beta_i$  es el coeficiente de regresión asociado al  $i$ -ésimo componente principal.
- $m$  es el número de componentes principales seleccionados para el modelo.

El proceso incluyó la transformación de los datos originales ( $X$ ) en componentes principales ( $Z$ ) utilizando la matriz de carga:

$$Z = X \cdot P$$

Posteriormente, el modelo de regresión se ajustó utilizando los componentes principales seleccionados ( $Z_1, Z_2, \dots, Z_m$ ) para explicar la variabilidad en el consumo de oxígeno ( $VO_2$ ). Las métricas de desempeño, como el error cuadrático medio ( $RMSE$ ) y el coeficiente de determinación ( $R^2$ ), se calcularon tanto en los conjuntos de entrenamiento como de prueba. Esto permitió evaluar la eficacia del modelo y validar su capacidad para capturar las relaciones clave entre las variables seleccionadas y el rendimiento físico.

El desempeño del modelo fue evaluado utilizando métricas clave como el error cuadrático medio ( $RMSE$ ) y el coeficiente de determinación ( $R^2$ ), tanto en los conjuntos de entrenamiento como de prueba. Estas métricas proporcionaron una medida del ajuste del modelo y de su capacidad predictiva.

El  $R^2$  en el conjunto de prueba fue de 0.89, lo que indica que el modelo explicó el 89% de la variabilidad observada en  $VO_2$ .

El error cuadrático medio ( $RMSE$ ) fue de 8.45, lo que refleja la precisión del modelo al predecir valores no observados.

### III. ANÁLISIS DE RESULTADOS

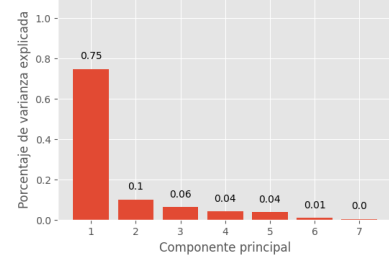
La interpretación de los resultados obtenidos a partir de los métodos aplicados reveló patrones significativos y conexiones relevantes entre las variables fisiológicas, demográficas y ambientales.

El análisis de componentes principales ( $PCA$ ) permitió identificar los principales factores que explican la variabilidad en los datos. Esto nos permitió reducir la dimensionalidad de los datos, preservando la mayor proporción de varianza posible en un número reducido de componentes. Este enfoque facilitó la identificación de patrones significativos en los conjuntos `test_measure` y `subject_info.csv`. En las variables fisiológicas, se observó que los primeros cuatro componentes principales explicaron más del 95% de la varianza total, destacando que variables como la frecuencia cardíaca ( $HR$ ), el consumo de oxígeno ( $VO_2$ ) y la ventilación pulmonar ( $VE$ ) fueron las de mayor influencia en los componentes principales. Por otro lado, en los datos demográficos y ambientales, los primeros cinco componentes principales explicaron una proporción similar de la varianza, con contribuciones significativas de la edad, el peso y la humedad. Como se evidenció en las Tablas 2, 3, 4 y 5.

#### a. Porcentaje de Varianza Explicada

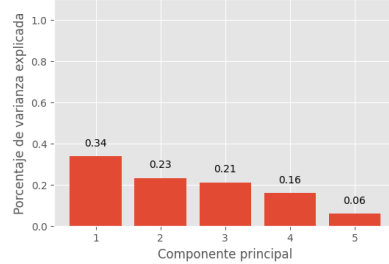
En `test_measure.csv`, el primer componente principal ( $PC1$ ) explicó el 74.7% de la varianza, mientras que los primeros cuatro componentes acumularon el 95% de la varianza total, según lo presentado en la Figura 1. En `subject_info.csv`, los primeros cinco componentes explicaron el 90% de la varianza total, destacando una mayor dispersión de la información entre múltiples factores demográficos y ambientales, tal como fue ilustrado en la Figura 2.

Porcentaje de varianza explicada por cada componente (test\_measure)



**Fig. 1:** Porcentaje de varianza explicada por cada componente principal en el conjunto `test_measure`.

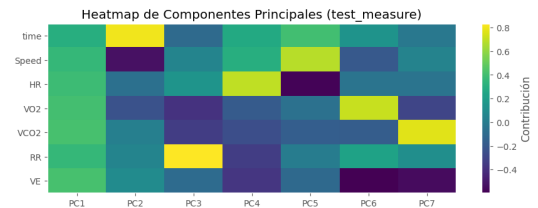
Porcentaje de varianza explicada por cada componente (subject\_info)



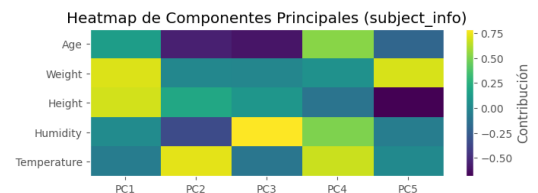
**Fig. 2:** Porcentaje de varianza explicada por cada componente principal en el conjunto `subject_info`.

#### b. Interpretación de las Cargas de los Componentes Principales

Las cargas de los componentes principales indicaron que las variables fisiológicas  $VO_2$ ,  $VE$  y  $HR$  tuvieron una contribución predominante en el primer componente principal de `test_measure.csv`, como quedó reflejado en la Figura 3. En `subject_info.csv`, la edad y el peso fueron las variables con mayor influencia en los componentes principales iniciales, tal como fue ilustrado en la Figura 4, lo que indica su impacto en las respuestas fisiológicas al ejercicio.



**Fig. 3:** Heatmap de Componentes Principales `test_measure`.



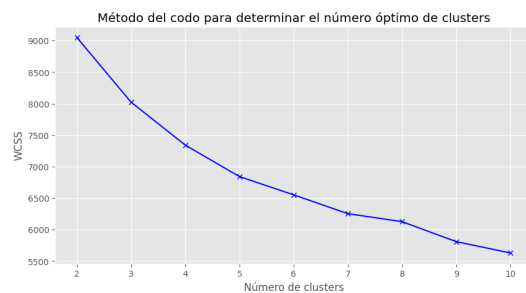
**Fig. 4:** Heatmap de Componentes Principales `subject_info`.

#### c. Agrupamiento de Participantes mediante Clustering

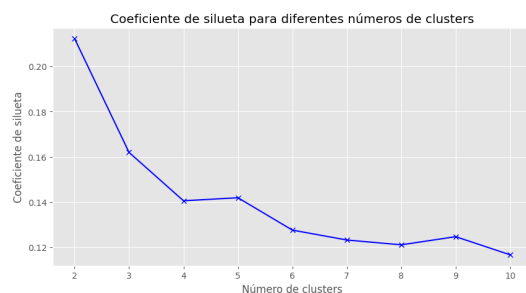
El algoritmo  $K$ -Means fue aplicado sobre las proyecciones obtenidas del  $PCA$  para identificar patrones subyacentes en

los datos. El método del codo y el coeficiente de silueta determinaron que tres *clusters* eran óptimos para ambos conjuntos de datos, como fue reportado en las Figuras 5, 6 y 7.

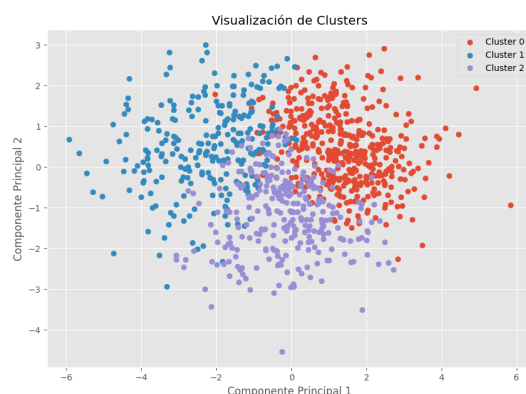
- **Cluster 1:** Participantes jóvenes con respuestas fisiológicas superiores al promedio y en condiciones ambientales favorables.
- **Cluster 2:** Participantes de mayor edad, con un rendimiento moderado y mayor influencia de la humedad en su respuesta.
- **Cluster 3:** Participantes con respuestas fisiológicas más bajas, posiblemente influenciadas por condiciones desfavorables de temperatura y peso elevado.



**Fig. 5:** Método del codo para determinar el número óptimo de clusters



**Fig. 6:** Coeficiente de silueta para diferentes números de clusters



**Fig. 7:** Visualización de Clusters

#### d. Análisis de Correlaciones entre Variables

Las correlaciones entre variables y *clusters* fueron analizadas para identificar relaciones significativas, tal como fue ilustrado en la Figura 8. Entre los hallazgos más relevantes:

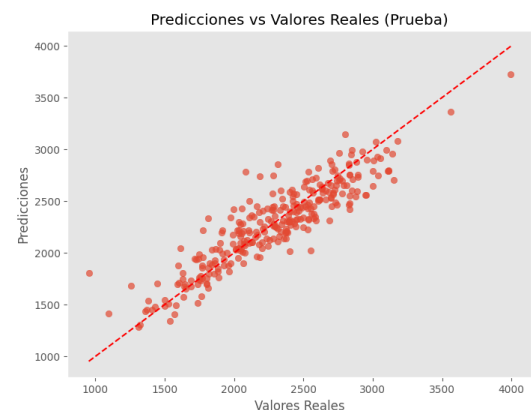
- VO2 presentó una correlación positiva significativa con VE ( $r = 0.89, p < 0.001$ ) y HR ( $r = 0.82, p < 0.001$ ).
- La temperatura ambiental mostró una correlación negativa moderada con el rendimiento ( $r = -0.45, p < 0.001$ ).



**Fig. 8:** Correlaciones entre Clusters y Variables

#### e. Modelo de Regresión Principal PCR

El modelo de regresión principal construido para predecir el consumo de oxígeno (VO2) en función de los componentes principales proporcionó un ajuste satisfactorio, con un  $R^2$  de 0.89 en el conjunto de prueba y un error cuadrático medio (RMSE) de 8.45. Estos resultados validaron la capacidad del modelo para capturar relaciones clave entre las variables y el rendimiento físico, como quedó reflejado en la Figura 9.



**Fig. 9:** Predicciones vs Valores Reales Prueba

Contrario a lo esperado, las variables ambientales como la temperatura y la humedad no influyeron de manera uniforme en las respuestas fisiológicas. Estos resultados pueden atribuirse a las condiciones controladas del experimento, que probablemente mitigaron su impacto. Esta observación resalta la necesidad de explorar estas relaciones bajo condiciones ambientales más variables.

## IV. CONCLUSIONES

Este estudio permitió abordar cómo los factores fisiológicos, demográficos y ambientales afectan el rendimiento en prue-





bas de ejercicio máximo. Mediante el uso de técnicas avanzadas como el análisis de componentes principales (*PCA*), el agrupamiento (*clustering*) y la regresión por componentes principales (*PCR*), se identificaron patrones significativos y relaciones clave entre las variables analizadas.

En primer lugar, el *PCA* demostró ser una herramienta efectiva bajo estas condiciones, ya que fue capaz de reducir la dimensionalidad de los datos mientras preservaba la mayor parte de la varianza. Entre las variables analizadas se incluyen factores fisiológicos, como el consumo de oxígeno ( $\text{VO}_2$ ), la ventilación pulmonar ( $\text{VE}$ ) y la frecuencia cardíaca ( $\text{HR}$ ), las cuales tuvieron una contribución notable al desempeño físico. Además, se evidenció que los factores demográficos, como la edad y el peso, también influyen de manera considerable en el rendimiento de los participantes; el junto con las condiciones ambientales, como la temperatura y la humedad, quedan en último lugar pero podría deberse al ambiente controlado en el experimento.

El análisis de agrupamiento reveló tres grupos principales entre los participantes, los cuales mostraron características distintivas. Esto destacó la forma en que las variables de los diferentes factores interactúan para definir perfiles segmentados. Por ejemplo, se encontró que los participantes con mejores respuestas fisiológicas al ejercicio se beneficiaron de condiciones ambientales más favorables y de un perfil demográfico más homogéneo.

Finalmente, la *PCR* resultó ser una herramienta poderosa para predecir el  $\text{VO}_2$  en función de los componentes principales obtenidos a través del *PCA*, logrando medidas de desempeño satisfactorias. Esto subraya la relevancia de incorporar múltiples variables en un marco analítico amplio para comprender plenamente el rendimiento físico.

Estos hallazgos tendrán implicaciones importantes para el diseño de programas de entrenamiento personalizados y más efectivos, optimizando el desempeño físico en diferentes poblaciones y condiciones ambientales.

## REFERENCES

- [1] T. R. Lopes, D. M. de Oliveira, L. A. Amoroso de Lima, and B. M. Silva, "Breathing variability during running in athletes: The role of sex, exercise intensity and breathing reserve," *Respiratory Physiology & Neurobiology*, vol. 331, p. 104350, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1569904824001435>
- [2] Mongin, Denis, G. Romero, Jeronimo, and J. R. A. Cruz, "Treadmill Maximal Exercise Tests from the Exercise Physiology and Human Performance Lab of the University of Malaga (version 1.0.1)," *PhysioNet*, 2021. [Online]. Available: <https://doi.org/10.13026/7ezk-j442>.
- [3] D. Mongin, C. Chabert, D. S. Courvoisier, J. García-Romero, and J. R. Alvero-Cruz, "Heart rate recovery to assess fitness: comparison of different calculation methods in a large cross-sectional study," *Research in Sports Medicine*, vol. 31, no. 2, pp. 157–170, 2023, PMID: 34308736. [Online]. Available: <https://doi.org/10.1080/15438627.2021.1954513>