

# Proyecto Final:

Estudio de Finalistas NBA Playoffs 2023-2024

---

*Fís. Guillermo Sierra Vargas*

*Fís. Isaias Siliceo Guzmán*



*Centro de Investigación en Matemáticas*

Maestría en Cómputo Estadístico

Estadística Multivariada

***Catedráticos:***

Dr. Rodrigo Macías Paez

Dr. Edgar Jiménez Peña

30 de octubre de 2024

# Tabla de contenidos

	Página
<b>1 Introducción</b>	<b>2</b>
1.1 Prueba de Bartlett	3
1.2 Rotación de Factores	4
1.2.1 Rotación varimax	4
1.3 Puntuaciones Factoriales	4
1.3.1 Método de Mínimos Cuadrados Ponderados	4
1.3.2 Método de Regresión	4
<b>2 Descripción de los datos</b>	<b>5</b>
<b>3 Análisis Exploratorio de los Datos (EDA)</b>	<b>6</b>
<b>4 Pruebas de Normalidad</b>	<b>10</b>
4.1 Prueba de normalidad de Henze-Zirkler	10
<b>5 Procedimiento</b>	<b>10</b>
5.1 Gráfico chi-cuadrada	11
<b>6 Análisis de Factores (FA)</b>	<b>12</b>
6.1 Método de componentes principales	12
6.2 Solución por Factor principal	13
6.3 Máxima verosimilitud	15
6.3.1 Rotación <i>varimax</i>	16
6.4 Prueba de Hipótesis	17
<b>7 Conclusión</b>	<b>20</b>
<b>A Tablas</b>	<b>21</b>
<b>B Referencias</b>	<b>21</b>

## 1. Introducción

El método de análisis de factores tiene sus raíces en los estudios psicológicos y estadísticos del siglo XIX y principios del XX. Este método fue desarrollado principalmente por dos pioneros en el campo de la estadística y la psicología: Karl Pearson y Charles Spearman.

Karl Pearson, un matemático y biólogo británico, es conocido por sus contribuciones fundamentales a la estadística. A finales del siglo XIX, Pearson desarrolló la técnica de análisis de componentes principales (PCA), un precursor importante del análisis factorial. El PCA se utiliza para reducir la dimensionalidad de un conjunto de datos mientras se preserva la mayor cantidad de variabilidad posible. Esta técnica es particularmente útil cuando se trabaja con grandes conjuntos de datos que contienen múltiples variables correlacionadas. Pearson introdujo la idea de que se podían encontrar combinaciones lineales de las variables originales que resumieran la información contenida en el conjunto de datos.

Charles Spearman, un psicólogo británico, es otro de los fundadores clave del análisis factorial. En 1904, Spearman publicó un artículo en el que introdujo el concepto de factor general (g). Spearman propuso que las habilidades cognitivas humanas podían explicarse en gran medida por un único factor subyacente, que llamó "g". Este factor general representaba la inteligencia general de una persona y se podía inferir a partir de varias pruebas cognitivas. Spearman desarrolló el análisis factorial para identificar y medir este factor general, así como otros factores específicos que podrían influir en diferentes habilidades cognitivas.

El análisis de factores se utiliza para identificar la estructura subyacente en un conjunto de datos. Es especialmente útil cuando se trabaja con grupos de variables que están correlacionadas entre sí. En la práctica, el análisis de factores ayuda a reducir la complejidad de los datos, facilitando la interpretación y la identificación de patrones significativos.

El modelo factorial permite explicar las relaciones entre variables observadas en términos de un número menor de variables no observadas llamadas factores. El modelo se puede expresar como:

$$\mathbf{X} = \mathbf{LF} + \epsilon, \quad (1)$$

donde:

- $\mathbf{X}$  es un vector de  $p \times 1$  de variables observadas,
- $\mathbf{L}$  es una matriz de  $p \times m$  de cargas factoriales,

- $\mathbf{F}$  es un vector de  $m \times 1$  de factores comunes,
- $\epsilon$  es un vector de  $p \times 1$  de factores únicos (errores),
- $p$  es el número de variables observadas,
- $m$  es el número de factores comunes.

Se asume que los factores  $\mathbf{F}$  no están correlacionados con los factores únicos  $\epsilon$  y que los factores únicos no están correlacionados entre sí.

Para construir dichos factores se pueden emplear dos metodos diferentes: un método algebraico basado en componentes principales y otro más estadístico basado en máxima verosimilitud. El método de componentes principales del análisis factorial tiene como objetivo reducir la dimensionalidad de los datos conservando la mayor variabilidad posible. Los componentes principales se obtienen resolviendo el problema de valores propios:

$$\mathbf{S}\mathbf{v}_i = l_i \mathbf{v}_i, \quad (2)$$

donde  $\mathbf{S}$  es la matriz de covarianza muestral,  $l_i$  son los valores propios, y  $\mathbf{v}_i$  son los vectores propios correspondientes. Las cargas factoriales  $\mathbf{L}$  se pueden estimar utilizando los primeros  $m$  vectores propios. Por otro lado el método de máxima verosimilitud estima las cargas factoriales y las varianzas únicas maximizando la función de verosimilitud de los datos observados bajo la suposición de normalidad multivariante. La función de log-verosimilitud se da por:

$$L(\mathbf{L}, \mathbf{\Psi}) = -\frac{N}{2} [p \log(2\pi) + \log |\mathbf{\Sigma}| + \text{tr}(\mathbf{\Sigma}^{-1} \mathbf{S})], \quad (3)$$

donde  $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^T + \mathbf{\Psi}$ ,  $\mathbf{\Psi}$  es una matriz diagonal de varianzas únicas, y  $N$  es el tamaño de la muestra.

### 1.1. Prueba de Bartlett

La prueba de Bartlett se utiliza para determinar el número de factores comunes en un modelo de análisis factorial. El estadístico de prueba se da por:

$$\chi^2 = - \left( N - 1 - \frac{2p + 5}{6} \right) \log \left( \frac{|\mathbf{S}|}{|\mathbf{\Sigma}|} \right), \quad (4)$$

donde  $\mathbf{S}$  es la matriz de covarianza muestral y  $\mathbf{\Sigma}$  es la matriz de covarianza estimada bajo la hipótesis nula. El estadístico de prueba sigue una distribución chi-cuadrado con grados de libertad:

$$\text{df} = \frac{1}{2} [p(p + 1) - 2mp - m(m - 1)]. \quad (5)$$

## 1.2. Rotación de Factores

Las rotaciones ortogonales tienen como objetivo hacer que la matriz de cargas factoriales  $\mathbf{L}$  sea más interpretable mientras se preserva la ortogonalidad de los factores. Los métodos comunes de rotación ortogonal incluyen Varimax, Quartimax y Equamax.

### 1.2.1. Rotación varimax

La rotación varimax se emplea para simplificar la representación de un subespacio específico, reduciéndolo a unos pocos elementos principales. Aunque el sistema de coordenadas permanece igual, lo que se ajusta es la base ortogonal para alinearse con esas coordenadas. El subespacio identificado a través del análisis de componentes principales o del análisis factorial suele estar representado por una base densa con muchos pesos diferentes, lo que complica su interpretación. La técnica varimax se denomina así porque maximiza la suma de las varianzas de las cargas cuadradas (las correlaciones cuadradas entre variables y factores).

## 1.3. Puntuaciones Factoriales

### 1.3.1. Método de Mínimos Cuadrados Ponderados

El método de mínimos cuadrados ponderados estima las puntuaciones factoriales minimizando la suma ponderada de las diferencias cuadráticas entre las variables observadas y sus valores estimados. Las puntuaciones factoriales  $\hat{\mathbf{f}}$  se obtienen resolviendo:

$$\hat{\mathbf{f}} = (\mathbf{L}^T \mathbf{W} \mathbf{L})^{-1} \mathbf{L}^T \mathbf{W} \mathbf{x}, \quad (6)$$

donde  $\mathbf{W}$  es una matriz de pesos.

### 1.3.2. Método de Regresión

El método de regresión estima las puntuaciones factoriales al ajustar las variables observadas a las cargas factoriales estimadas. Las puntuaciones factoriales  $\hat{\mathbf{f}}$  se dan por:

$$\hat{\mathbf{f}} = \mathbf{L}^T (\mathbf{L} \mathbf{L}^T + \mathbf{\Psi})^{-1} \mathbf{x}. \quad (7)$$

Este método es computacionalmente más simple y a menudo se utiliza en la práctica.

## 2. Descripción de los datos

Los datos correspondientes a los jugadores finalistas de los juegos tienen las siguientes

- Hablar acerca de cada una de las variables utilizadas en el o los datasets.
- Identificarlas como variables cuantitativas o cualitativas, etc.

La base de datos contiene 28 datos de los jugadores de dos equipos de baloncesto de la *National Basketball of Association (NBA)*, correspondientes a los jugadores de los dos equipos finalistas en los Playoffs 2023-2024. Los Celtics de Boston y Los Mavericks de Dallas. Las variables contenidas en esta base están descritas en el cuadro 1.

Variable	Descripción
Player	Indica el nombre del jugador
Team	Equipo del jugador Dallas o Boston
Pos	Posición del jugador en la cancha
Age	Edad del jugador en años
Ht	Altura del jugador en metros
Wt	Peso del jugador en kilogramos
Exp	Experiencia del jugador en años
Country	País de procedencia del jugador
PPG	Puntos por juego
RPG	Rebotes por juego
APG	Asistencias por juego
PIE	Estimador del impacto del jugador

Cuadro 1: Descripción general de las variables en la base de datos.

### 3. Análisis Exploratorio de los Datos (EDA)

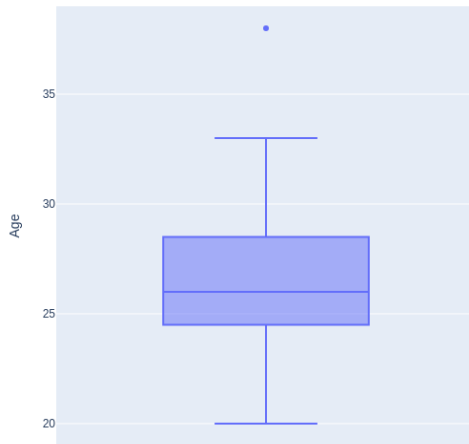
A continuación, en el Cuadro 2 se muestra un resumen de los estadísticos más relevantes de la base de datos, como lo es la media, desviación estándar, los mínimos y máximos de cada variable, así como los cuartiles de cada variable.

	Age (años)	Ht (m)	Wt (Kg)	Exp (años)	PPG	RPG	APG	PIE
<b>media</b>	26.68	2.01	98.54	5.39	10.82	4.29	2.37	9.76
<b>std</b>	4.29	0.08	8.94	3.98	8.19	2.13	2.06	3.32
<b>min</b>	20.00	1.85	81.00	0.00	1.70	1.20	0.50	5.50
<b>25 %</b>	24.75	1.96	92.75	2.75	5.08	3.12	1.00	7.68
<b>50 %</b>	26.00	2.01	96.00	5.00	8.60	3.70	1.60	8.95
<b>75 %</b>	28.25	2.06	106.75	7.00	13.28	5.52	3.02	11.38
<b>máx</b>	38.00	2.18	113.00	16.00	33.90	9.20	9.80	20.00

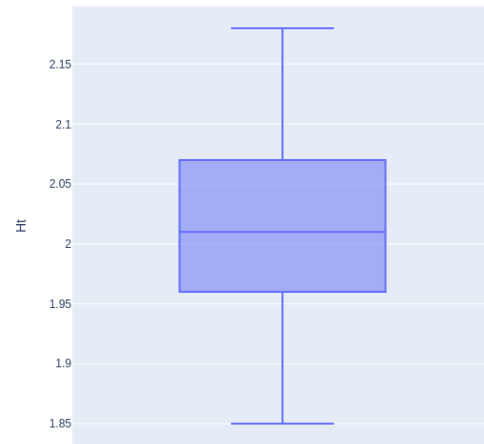
Cuadro 2: Resumen de las estadísticas de los jugadores de Dallas y Boston.

En las figuras 1, 2 y 3 se muestran los *boxplots* correspondientes a cada una de las variables. En algunos de las variables se encontraron datos atípicos, los cuales muestran a los jugadores estrella de cada equipo. En particular, el jugador más destacable es Luka Doncic, el cual tiene un impacto (PIE) considerablemente alto con respecto a los demás jugadores. En la figura 4 se muestra un mapa de calor de las correlaciones entre las variables cuantitativas de la base de datos. En este mapa se observa que variables como la experiencia del jugador (Exp) con su edad (Age) están altamente correlacionadas, el peso (Wt) y la altura (Ht) también tienen están correlacionados y finalmente las variables PPG, RPG, APG y PIE también muestran una correlación entre ellas.

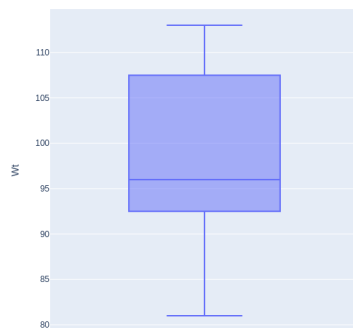
En los gráficos de dispersión múltiple mostrados en las figuras 5 se muestra una tendencia lineal entre las variables edad y experiencia; Peso y Altura; y las combinaciones posibles entre PIE, APG, RPG y PPG.



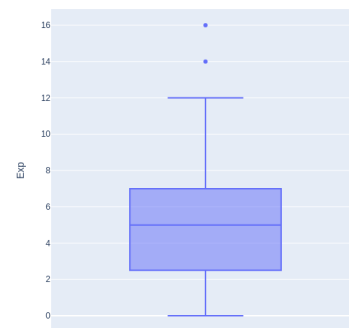
(a) El dato atípico corresponde al jugador de Boston, Al Horford, cuya edad es de 38 años.



(b) La altura no contiene datos atípicos.



(c) El peso de los jugadores no muestra datos atípicos.



(d) Los dos jugadores con más años de experiencia son Jrue Holiday con 14 años y Al Horford con 16 años. Ambos jugadores de Boston Celtics.

Figura 1: En esta figura se muestran las distribuciones de las variables correspondientes a (a) Edad y (b) Altura.

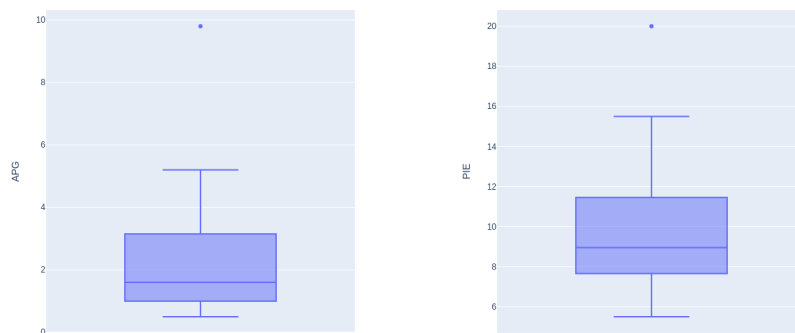




(a) Los dos jugadores con más puntos son Luka Doncic de Dallas, con 33.9 PPG y Jayson Tatum de Boston, con 26.9 PPG.

(b) Los rebotes por juego no tienen datos atípicos.

Figura 2: En estos Boxplots se muestran las distribuciones de las variables (a) Peso, (b) Experiencia, (c) Puntos por juego y (d) Rebotes por juego.



(a) El jugador con más asistencias es Luka Doncic de Dallas, con 9.8 APG.

(b) El jugador con más impacto es Luka Doncic con un PIE de 20.

Figura 3: En esta figura, se muestra la distribución de los datos, se observa un dato atípico en cada boxplot, correspondiente a Luka Doncic, jugador titular de Dallas.

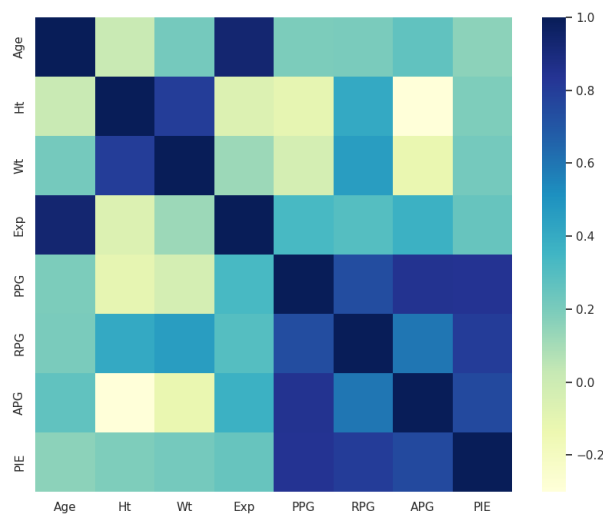


Figura 4: Mapa de calor que muestra la correlación entre las variables. En colores más oscuros se muestran aquellas variables que tienen una correlación positiva alta.

Diagrama de dispersión múltiple

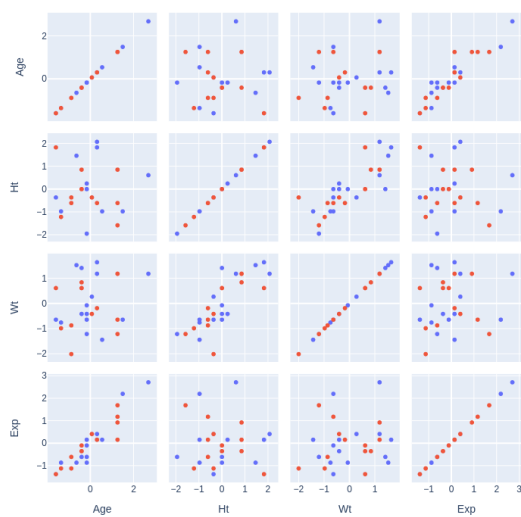


Diagrama de dispersión múltiple

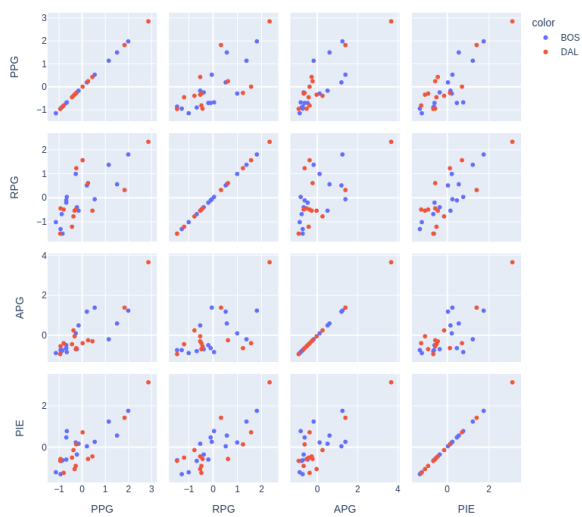


Figura 5: Gráficos de dispersión múltiple correspondientes a las zonas más oscuras del mapa de calor que describe la correlación de las variables en la figura 4.

## 4. Pruebas de Normalidad

### 4.1. Prueba de normalidad de Henze-Zirkler

La prueba de Henze-Zirkler se define a partir de la siguiente estadística de prueba:

$$HZ = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \left(1 + \frac{\gamma}{2} D_{ij}^2\right)^{-\frac{p}{2}} - 2 \left(1 + \frac{\gamma}{2}\right)^{-\frac{p}{2}} + (1 + \gamma)^{-\frac{p}{2}}, \quad (8)$$

donde:

- $n$  es el tamaño de la muestra.
- $p$  es la dimensión de los datos.
- $D_{ij}$  es la distancia de Mahalanobis entre las observaciones  $i$  y  $j$ .
- $\gamma$  es un parámetro positivo que se recomienda elegir como  $\gamma = 1.0$ .

La distancia de Mahalanobis  $D_{ij}$  se calcula como:

$$D_{ij} = \sqrt{(x_i - x_j)^T S^{-1} (x_i - x_j)}, \quad (9)$$

donde  $x_i$  y  $x_j$  son las observaciones  $i$  y  $j$ , y  $S$  es la matriz de covarianza muestral.

## 5. Procedimiento

El procedimiento para llevar a cabo la prueba de Henze-Zirkler es el siguiente:

1. Calcular la distancia de Mahalanobis  $D_{ij}$  para todas las parejas de observaciones.
2. Sustituir  $D_{ij}$  en la fórmula de la estadística  $HZ$ .
3. Comparar el valor de  $HZ$  con los valores críticos de la distribución asintótica del test o realizar una simulación para obtener el valor  $p$ .

La prueba de Henze-Zirkler es una herramienta poderosa para la evaluación de la normalidad multivariada en conjuntos de datos. Su formulación, basada en la distancia de Mahalanobis y funciones características, permite una evaluación robusta y efectiva de esta suposición crucial en análisis multivariados.

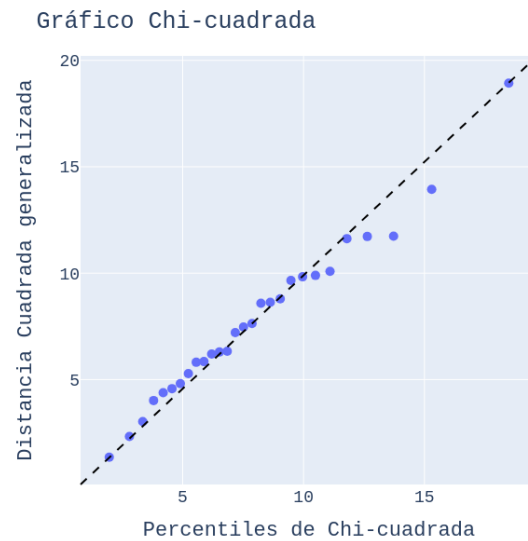


Figura 6: En esta figura se muestran las distancias cuadradas generalizadas de los datos ordenados en función de los percentiles  $q_{c,8}((j - 1/2)/28)$  de una distribución  $\chi^2$

### 5.1. Gráfico chi-cuadrada

A continuación, se verifica la normalidad de los datos mediante la visualización de un gráfico Chi-cuadrado. En este gráfico se muestran

## 6. Análisis de Factores (FA)

### 6.1. Método de componentes principales

Calculando los valores propios de la matriz de correlación  $R$  de los datos,

$$\text{eigenvalues} = (3.669, 1.637, 2.064, 0.224, 0.169, 0.048, 0.083, 0.106)$$

y siguiendo el criterio de Kaiser, se encuentra que el número de factores apropiado para describir los datos es  $m = 3$ , ya que hay 3 valores propios mayores que 1. Además, en la Figura 7 se muestra el gráfico de codo y la varianza acumulada de cada una de las componentes. Se puede observar en esta figura que bastan 3 componentes para describir a los datos con un 90 % de la varianza original.

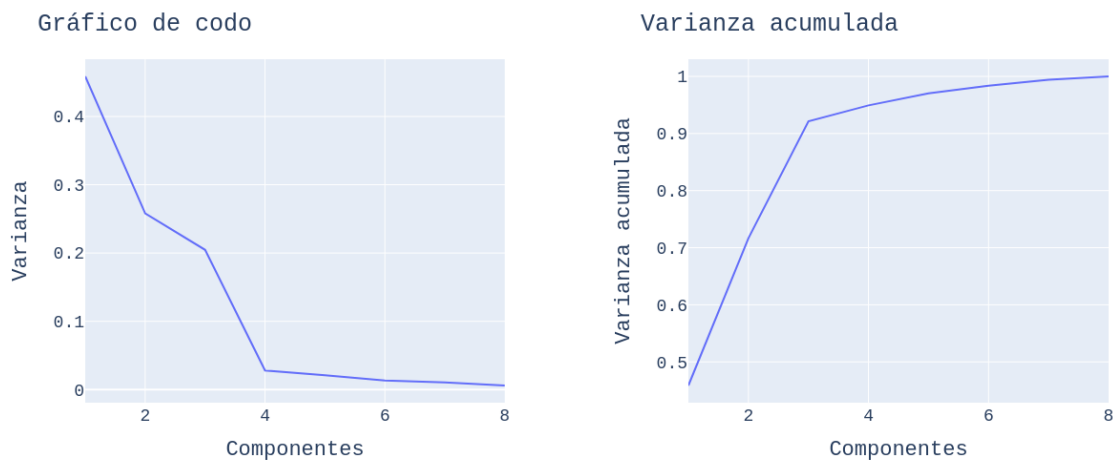


Figura 7: Gráfico de codo y varianza acumulada de las primeras componentes del análisis de factores. se puede observar que  $m = 3$  factores superan el 90 % de varianza acumulada en los datos. Lo cual indica que un gran porcentaje de la varianza se encuentra en las primeras 3 componentes principales.

Aplicando el método por componentes principales, determinamos la matriz de cargas  $L$ , la cual se presenta en el cuadro 3. En amarillo se muestran aquellos valores con mayor aporte a cada factor. Podemos concluir que los factores sin rotación no tienen una interpretación sencilla, ya que se combinan conceptos como el de la altura con el impacto en el juego o la experiencia y el número de rebotes por juego. De modo que continuaremos el estudio haciendo una rotación de estas cargas.

Variable	F1	F2	F3	Comunalidades $h_i^2$	Varianzas Específicas $\psi_i$
Age	0.257	0.074	0.146	0.09	0.91
Ht	0.004	0.659	0.622	0.82	0.18
Wt	0.668	-0.061	0.071	0.45	0.55
Exp	-0.009	-0.429	0.670	0.63	0.37
PPG	-0.175	-0.230	0.025	0.08	0.92
RPG	0.026	0.254	-0.347	0.19	0.81
APG	0.160	0.477	-0.101	0.26	0.74
PIE	0.657	-0.164	-0.079	0.46	0.54
Eigenvalues	3.669	2.064	1.637		
Var. Acumulada	0.4586	0.7166	0.9212		

Cuadro 3: Tabla de los primeros 3 factores correspondientes a la aplicación del método por componentes principales. Con sus comunalidades y varianzas específicas.

## 6.2. Solución por Factor principal

Aplicando una rotación a la solución por componentes principales hasta obtener un error  $\epsilon = 0.01$ , se obtiene una matriz de cargas dada por el Cuadro 4. En esta tabla es posible identificar que los factores tienen una interpretación. El primer factor corresponde a todos los indicadores de desempeño en la cancha del jugador. En el segundo factor están las dos variables que describen la talla de los jugadores. Finalmente, el tercer factor describe la experiencia de los jugadores.

Ya que los factores tienen una interpretación, se procede a realizar el cálculo de los *factor scores*. Los resultados de la proyección en los primeros dos factores se muestran en las figuras 8. En el primer gráfico, se muestra de manera ilustrativa cómo la talla del jugador influye en la posición que toman en el juego. Los jugadores con una talla mayor, están en la parte superior del gráfico y corresponden a jugadores defensivos. así mismo, las posiciones ofensivas se caracterizan por tener una talla menor para una mayor agilidad. Por otro lado, en el segundo gráfico, se muestra, de izquierda a derecha, el desempeño en la cancha de los jugadores. En este gráfico es posible observar que los jugadores en azul (Celtics) tienen un mejor desempeño en la cancha ya que hay 7 jugadores a la derecha del origen. Dallas tiene apenas 4 jugadores a la derecha del origen, de los cuales destaca Luka Doncic y Kyrie Irving, los jugadores más destacados de Dallas.

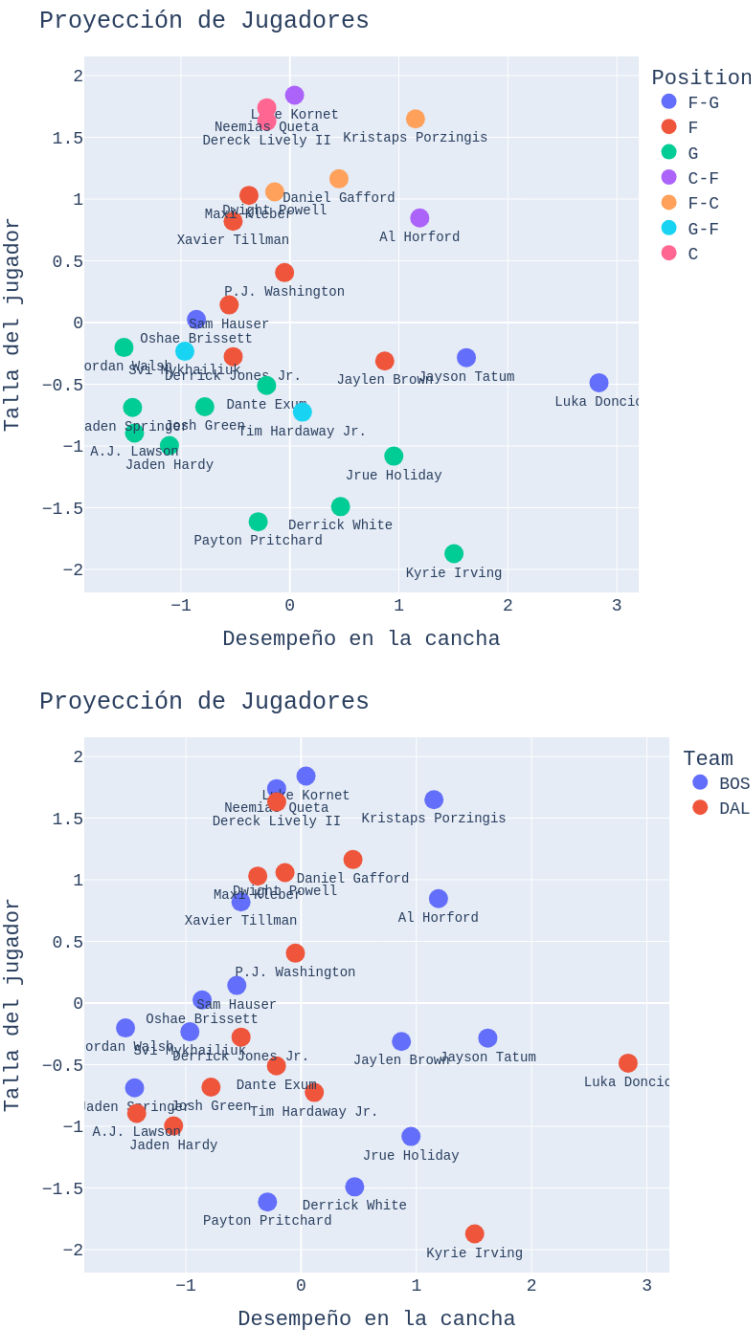


Figura 8: Proyección de los jugadores (*factor scores*) asociados a las cargas del Cuadro 4

Variable	F1	F2	F3	Comunalidades $h_i^2$	Varianzas Específicas $\psi_i$
Age	0.497	-0.003	0.833	0.93	0.07
Ht	0.134	0.931	-0.057	0.87	0.13
Wt	0.262	0.825	0.092	0.76	0.24
Exp	0.589	-0.106	0.757	0.93	0.07
PPG	0.869	-0.258	-0.281	0.90	0.10
RPG	0.844	0.302	-0.243	0.86	0.14
APG	0.806	-0.419	-0.173	0.86	0.14
PIE	0.865	0.034	-0.332	0.86	0.14
Eigenvalues	3.669	2.064	1.637		
Var. Acumulada	0.4586	0.7166	0.9212		

Cuadro 4: Tabla de los primeros 3 factores correspondientes a la rotación por el método del factor principal. Se presentan las comunalidades y varianzas específicas. En esta tabla se puede observar que con la rotación, los factores son interpretables.

### 6.3. Máxima verosimilitud

En esta sección se aborda el análisis de factores por el método de máxima verosimilitud. En el Cuadro 5 se muestra el resultado de la estimación de las cargas por el método de Máxima Verosimilitud sin aplicar una rotación. Los factores no tienen una interpretación directa, ya que combinan conceptos de nuevamente como la talla del jugador y su número de asistencias por juego.

Variable	F1	F2	F3	Comunalidades $h_i^2$	Varianzas Específicas $\psi_i$
Age	0.868	-0.348	-0.281	0.95	0.05
Ht	0.024	0.615	-0.748	0.94	0.06
Wt	0.215	0.487	-0.667	0.73	0.27
Exp	0.907	-0.298	-0.135	0.93	0.07
PPG	0.602	0.482	0.554	0.90	0.10
RPG	0.560	0.731	0.073	0.85	0.15
APG	0.614	0.278	0.635	0.86	0.14
PIE	0.553	0.671	0.329	0.86	0.14
Var. Acumulada	0.3727	0.6377	0.8784		

Cuadro 5: Cargas determinadas por el método de máxima verosimilitud sin aplicar una rotación. Se puede observar que no hay una interpretación clara utilizando estas cargas.



### 6.3.1. Rotación *varimax*

Se procedió a realizar nuevamente el procedimiento pero esta vez con una rotación *varimax*, los resultados se muestran en el Cuadro 6. Se puede observar que las proyecciones (9 y 10) más ilustrativas se observan para los factores 2 y 3. Los cuales describen nuevamente el desempeño de los jugadores de izquierda (menor desempeño) a derecha (un desempeño destacable). Y de abajo (Jugadores con menor talla) hacia arriba (jugadores de mayor talla). En la figura 9 se observa algo similar a lo que se observó en la figura 8, en donde, hay más jugadores del equipo de Boston a la derecha del origen, lo cual muestra una desventaja considerable para los jugadores de Dallas.

Variable	F1	F2	F3	Comunalidades $h_i^2$	Varianzas Específicas $\psi_i$
Age	0.971	0.078	-0.066	0.95	0.05
Ht	-0.049	-0.005	-0.968	0.94	0.06
Wt	0.149	0.058	-0.839	0.73	0.27
Exp	0.940	0.212	0.015	0.93	0.07
PPG	0.136	0.935	0.096	0.90	0.10
RPG	0.122	0.807	-0.432	0.85	0.15
APG	0.216	0.854	0.284	0.86	0.14
PIE	0.072	0.906	-0.194	0.86	0.14
Var. Acumulada	0.2422	0.6336	0.8784		

Cuadro 6: Cargas determinadas por el método de máxima verosimilitud con una rotación *varimax*. La interpretación de estos factores es directa, ya que el factor 1 (F1) recupera información de la Experiencia, el factor 2 (F2) denota el desempeño en la cancha y el factor 3 (F3) la talla del jugador. Además, se observa que el porcentaje de la varianza original es del 87 %.

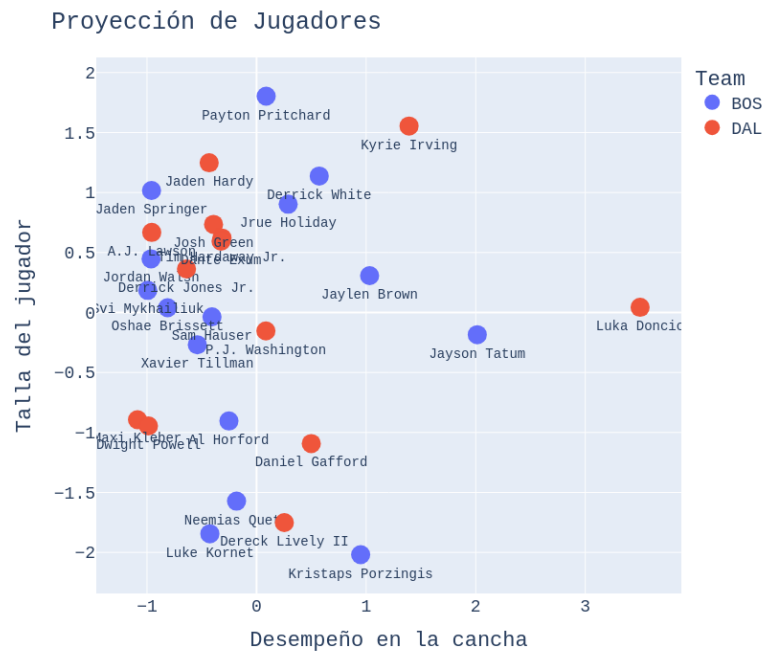


Figura 9: Proyección de los jugadores (*factor scores*) asociados a las cargas del Cuadro 6

## 6.4. Prueba de Hipótesis

Se establece la hipótesis nula,

$$H_0 : \Sigma = LL^T + \Psi \quad (10)$$

y la hipótesis alternativa

$$H_1 : \Sigma, \text{ Alguna otra matriz positiva definida} \quad (11)$$

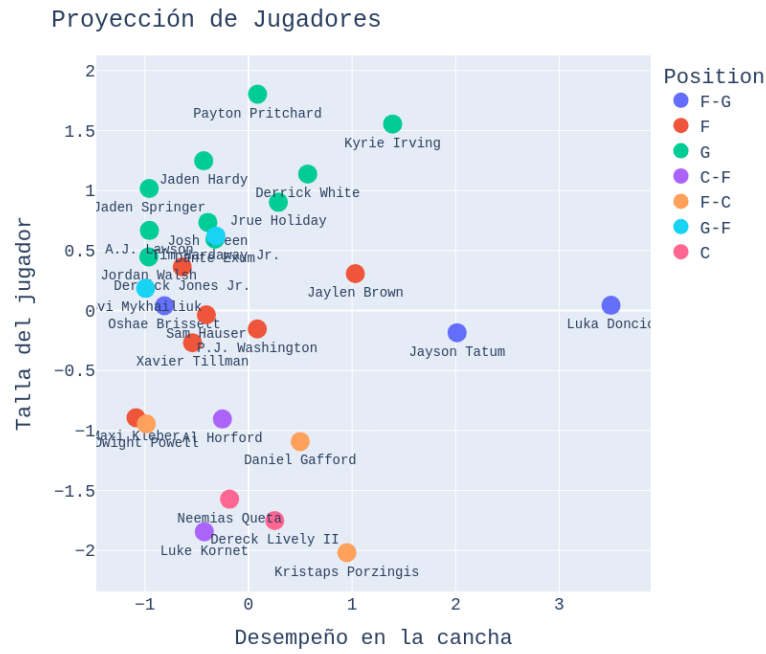


Figura 10: Proyección de los jugadores (*factor scores*) asociados a las cargas del Cuadro 6

Ya que el número de grados de libertad de la distribución  $\chi^2$  debe ser positivo, debe cumplirse, para  $p = 8$  que

$$\begin{aligned}
 m &\leq \frac{1}{2} \left( 2p + 1 - \sqrt{8p + 1} \right) \\
 m &\leq \frac{1}{2} \left( 2(8) + 1 - \sqrt{8(8) + 1} \right) \\
 m &\leq \frac{1}{2} \left( 17 - \sqrt{65} \right) \\
 m &\leq \frac{1}{2} (8.9377) \\
 m &\leq 4.4688
 \end{aligned}$$

De modo que el test es válido para  $m = 3$  factores. Se determinan los grados de libertad

$$\text{dof} = \frac{1}{2}[(p - m)^2 - p - m]/2$$

$$\text{dof} = \frac{1}{2}[(8 - 3)^2 - 8 - 3]/2$$

$$\text{dof} = \frac{1}{2}[25 - 8 - 3]/2$$

$$\text{dof} = \frac{1}{2}[14]/2$$

$$\text{dof} = 7$$

Utilizando la corrección de Barlett, se rechaza  $H_0$  en un nivel  $\alpha = 0.05$  de significancia si

$$(n - 1 - (2p + 4m + 5)/6) \ln \frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}|}{|S_n|} > \chi^2_{\frac{1}{2}[(p-m)^2 - p - m]/2}(\alpha)$$

$$(28 - 1 - (2(8) + 4(3) + 5)/6) \ln \frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}|}{|S_n|} > \chi^2_7(0.05)$$

$$(21.5)(0.020182) > 14.0671$$

$$0.433925 > 14.0671$$

De modo que fallamos en rechazar la hipótesis nula con un grado de significancia de  $\alpha = 0.05$  dado que

$$0.433925 < 14.0671$$

Por lo tanto, la matriz de covarianzas se puede aproximar mediante

$$\Sigma = LL^T + \Psi$$

## 7. Conclusión

Los resultados obtenidos describen los datos mediante tres variables latentes de manera exitosa cuando se aplican ambos métodos de rotación. El porcentaje de varianza original recuperado por estos factores supera el 80 %, de modo que se trata una buena aproximación de la matriz de covarianzas. Además, no se rechaza la hipótesis de que los datos provengan de una distribución normal multivariada.

El análisis de factores por el método de factor principal arroja una matriz de cargas que se presenta en el Cuadro 4, en este se observa en color amarillo que los factores se cargan en variables que tienen una alta correlación entre ellas, lo cual se ve reflejado en el comportamiento que exhiben los *factor scores* en la figura 8, ya que destacan características esperadas de los jugadores. Por otro lado, para el método de Máxima verosimilitud con una rotación *varimax*, se obtiene una matriz de cargas igualmente interpretable que con el método anterior. Con este método las cargas describen los mismos rasgos que con el primero. Estos resultados se observan en el cuadro 6 y las figuras 9 y 10.

Las proyecciones en baja dimensión de los datos revelan que hay un desbalance considerable entre los equipos, por una parte, Boston cuenta con una ventaja en cuanto al número de jugadores con buenas estadísticas de juego, los cuales están bien diversificados en las diferentes posiciones de la cancha. En cambio, el equipo de Dallas, sólo cuenta con dos jugadores destacables, lo cual podría revelar que Boston será un contrincante difícil de superar durante las finales de los Playoffs 2023-2024. Así mismo, la distribución de los datos en el gráfico por posiciones muestra cómo la talla de cada jugador influye en las posiciones de juego. Lo cual podría utilizarse para introducir nuevos jugadores en alguna posición que esté de acuerdo a sus características físicas o desempeño en la cancha.

Estos datos se recabaron previo al primer juego de las finales, el cual se llevó acabo el 6 de junio de 2024 y el segundo juego que se llevó acabo el 9 de junio de 2024. Los resultados de ambos juegos se muestran en la figura 11. La victoria en los primeros 2 juegos es para Celtics. Así mismo, el desbalance en el desempeño de los jugadores en la cancha se ve reflejado en los puntos por jugadore en la figura 12, lo cual se observa en los datos proyectados en baja dimensión.

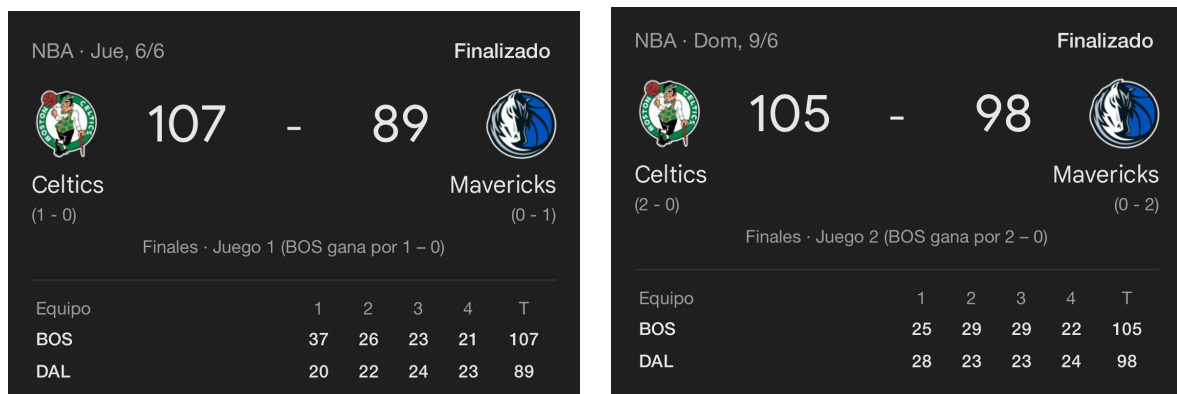


Figura 11: Resultados del marcador del primer y segundo juego de la final de Boston vs. Dallas en los PlayOffs 2023-2024

## A. Tablas

En la tabla 7 se muestran los datos que se utilizaron para elaborar el análisis de factores en este proyecto.

## B. Referencias

### Referencias

- [1] Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. Prentice Hall.
- [2] <https://www.sportingnews.com/mx/nba?gr=www>
- [3] <https://www.britannica.com/biography/Karl-Pearson>
- [4] Thomson, Godfrey (1947). "Charles Spearman. 1863-1945". *Obituary Notices of Fellows of the Royal Society*, 5(15): 373–385. doi:10.1098/rsbm.1947.0006

Celtics		Mavericks			
Jugador		Min	Reb	Ast	Pts
0	J. Tatum · A	42	11	5	16
7	J. Brown · A	37	6	2	22
42	A. Horford · P	30	7	3	10
4	J. Holiday · B	35	8	5	12
9	D. White · B	35	2	5	15
8	K. PORZINGIS	21	6	0	20
30	S. Hauser	16	4	1	8
11	P. Pritchard	15	2	2	0
12	O. Brissett	3	0	0	0
40	L. Kornet	3	1	0	2
50	S. Mykhailiuk	3	0	0	2

Celtics		Mavericks			
Jugador		Min	Reb	Ast	Pts
25	P. Washington · A	36	8	1	14
55	D. Jones Jr. · A	29	6	0	5
21	D. Gafford · P	14	3	1	8
77	L. Doncic · B	38	10	1	30
11	K. Irving · B	36	3	2	12
8	J. Green	20	3	0	3
42	M. Kleber	19	0	1	2
2	D. Lively II	18	5	1	2
1	J. Hardy	11	3	1	13
10	T. Hardaway Jr.	8	0	0	0
0	D. EXUM	5	2	1	0

Figura 12: Estadísticas de cada jugador en el primer juego de las finales de los PPlayoffs 2023-2024. En estas dos tablas se puede observar que el número de puntos por jugador por parte de Boston muestra un balance. Por otro lado, la diferencia de puntos que tienen jugadores como Luka Doncic y Kyrie Irving con respecto a sus compañeros es considerable. Lo cual se ve reflejado en la proyección que se elaboró con Análisis de factores en este proyecto.

Player	Team	Pos	Age	Ht	Wt	Exp	Country	PPG	RPG	APG	PIE
Jayson Tatum	BOS	F-G	26	2.03	95	6	USA	26.9	8.1	4.9	15.5
Jaylen Brown	BOS	F	27	1.98	101	7	USA	23.0	5.5	3.6	11.6
Derrick White	BOS	G	29	1.93	86	6	USA	15.2	4.2	5.2	10.6
Jrue Holiday	BOS	G	33	1.93	93	14	USA	12.5	5.4	4.8	9.9
Al Horford	BOS	C-F	38	2.06	109	16	DOM	8.6	6.4	2.6	10.5
Payton Pritchard	BOS	G	26	1.85	88	3	USA	9.6	3.2	3.4	10.3
Sam Hauser	BOS	F	26	2.01	98	2	USA	9.0	3.5	1.0	8.6
Kristaps Porzingis	BOS	F-C	28	2.18	109	7	LAT	20.1	7.2	2.0	13.8
Luke Kornet	BOS	C-F	28	2.16	113	6	USA	5.3	4.1	1.1	11.3
Oshae Brissett	BOS	F-G	25	2.01	95	4	CAN	3.7	2.9	0.8	7.6
Xavier Tillman	BOS	F	25	2.01	111	3	USA	5.3	3.9	1.4	7.8
Jaden Springer	BOS	G	21	1.93	92	2	USA	3.3	1.6	0.9	5.5
Svi Mykhailiuk	BOS	G-F	26	2.01	93	5	UKR	4.0	1.2	0.9	7.7
Jordan Walsh	BOS	G	20	1.98	93	0	USA	1.7	2.2	0.6	5.8
Neemias Queta	BOS	C	24	2.13	112	2	POR	5.5	4.4	0.7	12.3
Luka Doncic	DAL	F-G	25	2.01	104	5	SLV	33.9	9.2	9.8	20.0
Kyrie Irving	DAL	G	32	1.88	88	12	AUS	25.6	5.0	5.2	14.4
P.J. Washington	DAL	F	25	2.01	104	4	USA	12.9	5.6	1.9	7.9
Derrick Jones Jr.	DAL	F	27	1.98	95	7	USA	8.6	3.3	1.0	6.8
Daniel Gafford	DAL	F-C	25	2.08	106	4	USA	11.0	7.6	1.6	12.1
Dereck Lively II	DAL	C	20	2.16	104	0	USA	8.8	6.9	1.1	10.2
Josh Green	DAL	G	23	1.96	91	3	AUS	8.2	3.2	2.3	6.3
Jaden Hardy	DAL	G	21	1.91	90	1	USA	7.3	1.8	1.5	8.1
Tim Hardaway Jr.	DAL	G-F	32	1.96	93	10	USA	14.4	3.2	1.8	8.3
Maxi Kleber	DAL	F	32	2.08	109	6	GER	4.4	3.3	1.6	5.7
Dante Exum	DAL	G	28	1.96	97	6	AUS	7.8	2.7	2.9	9.3
A.J. Lawson	DAL	G	23	1.98	81	1	CAN	3.2	1.2	0.5	7.6
Dwight Powell	DAL	F-C	32	2.08	109	9	CAN	3.3	3.4	1.3	7.9

Cuadro 7: Datos utilizados para el análisis de factores en este proyecto.