

Analisis Estadístico sobre las finales de la NBA

Análisis de Factores de los jugadores finalistas de los Playoffs 2023-2024

Equipo:

Fís. Guillermo Sierra Vargas

Fís. Isaías Siliceo Guzmán

Estadística Multivariada

Centro de Investigación en Matemáticas (CIMAT-MTY)



CIMAT



CONAHCYT

CONSEJO NACIONAL DE HUMANIDADES
CIENCIAS Y TECNOLOGÍAS

CONTENIDO

1. Introducción

2. Descripción de los datos

3. Análisis de factores

4. Conclusiones

5. Referencias

Objetivos:

1. Realizar un análisis descriptivo
2. Implementar un método de reducción de dimensionalidad (Análisis de Factores)
3. Interpretar los factores
4. Mostrar insights obtenidos

Introducción

PLAYOFFS

- Torneo anual de Baloncesto organizado por la NBA en USA.
- Se selecciona al ganador de la liga.
- Mavericks (Dallas) Vs Celtics (Boston)



Figura: Equipos finalistas





Figura: Bracket oficial. Fuente: NBA.com

DATOS

Players ▾ Bios ▾

SEASON 2023-24 ▾	SEASON TYPE Playoffs ▾	PER MODE Per Game ▾	SEASON SEGMENT Last Game ▾
Player Bio		Game Situation	
POSITION All Positions ▾	TEAM Boston Celtics ▾	OUTCOME All Outcomes ▾	DATE RANGE
STARTER/BENCH All Players ▾	VS TEAM VS Dallas Mavericks ▾	LOCATION All Locations ▾	PLAYOFF ROUND All Playoff Rounds ▾
EXPERIENCE All Experience ▾	DIVISION All Divisions ▾	SHOT CLOCK RANGE All Shot Clock Ranges ▾	DATE FROM MM/DD/YYYY 📅
DRAFT YEAR All Draft Years ▾	VS DIVISION VS All Divisions ▾	QUARTER All Quarters ▾	DATE TO MM/DD/YYYY 📅
DRAFT PICK All Draft Rounds ▾	CONFERENCE All Conferences ▾	BY HALF Entire Game ▾	
COLLEGE All Affiliations ▾	VS CONFERENCE VS All Conferences ▾		
COUNTRY All Countries ▾	IN-SEASON TOURNAMENT ROUND ... ▾		
HEIGHT All Heights ▾			
WEIGHT All Weights ▾			

Reset Filters **Get Stats**

[Advanced Filters](#) ^

Team: Boston Celtics × VS Team: VS Dallas Mavericks × Last N Games: Last Game ×

Figura: Portal de datos oficial: NBA bios

PRUEBAS DE HIPÓTESIS

Se establece la hipótesis nula,

$$H_0 : \Sigma = LL^T + \Psi \quad (1)$$

y la hipótesis alternativa

$$H_1 : \Sigma, \text{ Alguna otra matriz positiva definida} \quad (2)$$

PRUEBAS DE HIPÓTESIS

El número de grados de libertad de la distribución χ^2 debe ser positivo, debe cumplirse para $p = 8$ que

$$m \leq \frac{1}{2} \left(2p + 1 - \sqrt{8p + 1} \right)$$

$$m \leq \frac{1}{2} \left(2(8) + 1 - \sqrt{8(8) + 1} \right)$$

$$m \leq 4.4688$$

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

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

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

$$\text{dof} = 7$$

PRUEBAS DE HPÓTESIS

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

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

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

$$0.433925 > 14.0671 \rightarrow \text{contradicción}$$

De modo que fallamos en rechazar la hipótesis nula. Por lo tanto, la matriz de covarianzas se puede aproximar mediante $\Sigma = LL^T + \Psi$.

PRUEBAS DE HIPÓTESIS

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.

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

$$HZ = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \left(1 + \frac{\gamma}{2} D_{ij}^2\right)^{-\frac{p}{2}} - 2 \left(1 + \frac{\gamma}{2}\right)^{-\frac{p}{2}} + (1 + \gamma)^{-\frac{p}{2}}, \quad (3)$$

- 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.

Descripción de los datos

ANÁLISIS EXPLORATORIO

- La base de datos contiene información correspondientes a jugadores de dos equipos de baloncesto de la *National Basketball of Association (NBA)*.
- Está conformada por 28 observaciones (jugadores) y 12 variables: 1 cadena de texto, 3 categóricas, 8 numéricas.
- Los equipos analizados son los finalistas en los Playoffs 2023-2024. Los Celtics de Boston y Los Mavericks de Dallas.

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

ANÁLISIS EXPLORATORIO

- Se calculó la media, desviación estándar, valores extremos y cuartiles para las 8 variables numéricas.

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

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

ANÁLISIS EXPLORATORIO

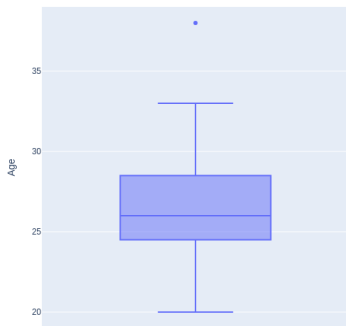


Figura: El dato atípico corresponde al jugador de Boston, Al Horford, cuya edad es de 38 años.

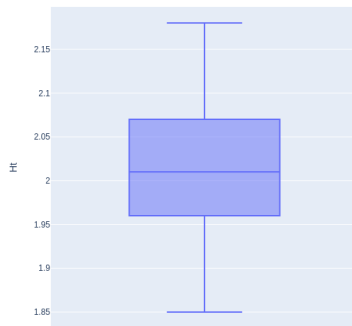


Figura: La altura no contiene datos atípicos.

ANÁLISIS EXPLORATORIO

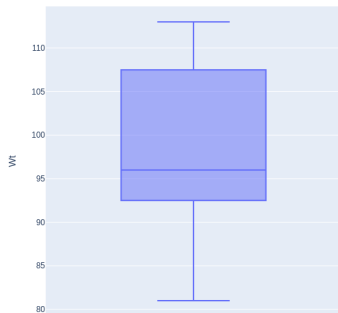


Figura: El peso de los jugadores no muestra datos atípicos.

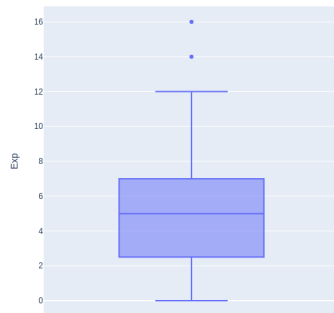


Figura: Los jugadores más experimentados son Jrue Holiday con 14 años y Al Horford con 16 años. Ambos de Boston Celtics.

ANÁLISIS EXPLORATORIO

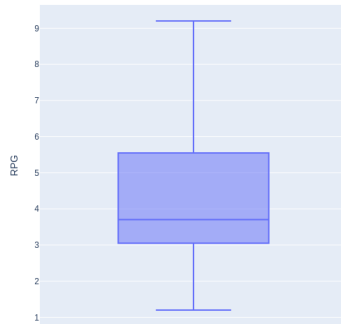
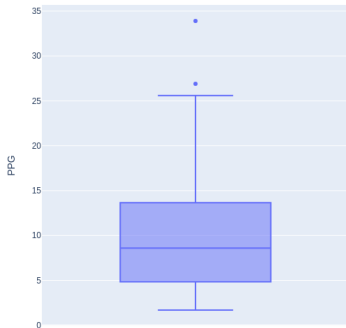


Figura: Los jugadores con más puntos son Luka Doncic de Dallas, con 33.9 PPG y Jayson Tatum de Boston, con 26.9 PPG.

Figura: Los rebotes por juego no tienen datos atípicos.

LÍDERES DE EQUIPO

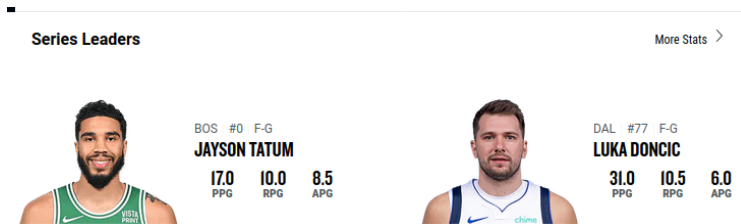


Figura: Outliers: Jayson Tatum, y Luka Doncic. Consultado 12/06/2024. Fuente: NBA finals

ANÁLISIS EXPLORATORIO

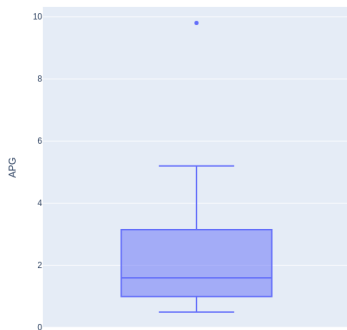


Figura: El jugador con más asistencias es Luka Doncic de Dallas, con 9.8 APG.

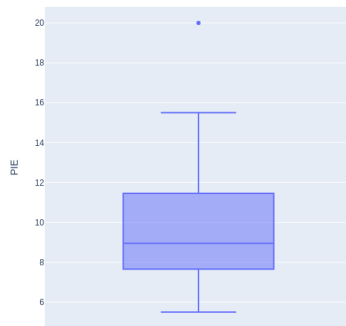


Figura: El jugador con más impacto es Luka Doncic con un PIE de 20.

ANÁLISIS EXPLORATORIO

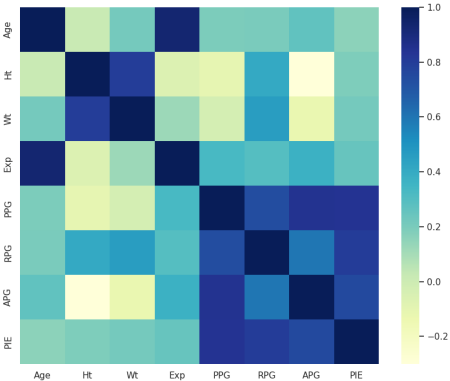


Figura: Mapa de calor que muestra la correlación entre las variables. En colores más oscuros se muestran aquellas variables que tienen una correlación positiva alta.

ANÁLISIS EXPLORATORIO

Diagrama de dispersión múltiple

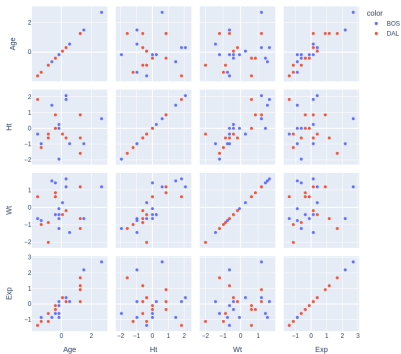


Diagrama de dispersión múltiple

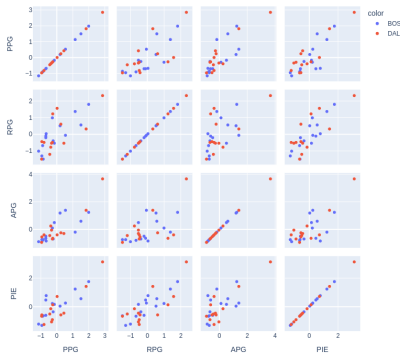


Figura: Gráfico de dispersión múltiple correspondiente a las zonas más oscuras del mapa de calor.

Figura: Gráfico de dispersión múltiple correspondiente a las zonas más oscuras del mapa de calor.

PRUEBAS DE NORMALIDAD

- La prueba de Henze-Zirkler arroja:

$$p\text{-valor} = 0.05821$$

De modo que se falla en rechazar la hipótesis $H_0 : X \sim \mathcal{N}(\Sigma, \mu)$

- Además, el gráfico Chi-Cuadrada aparentemente es normal.

Gráfico Chi-cuadrada

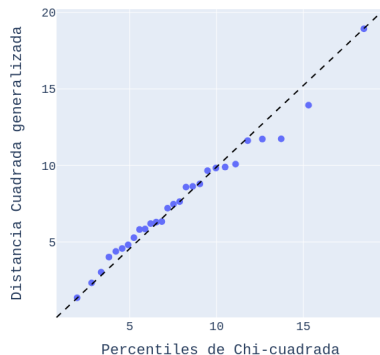


Figura: La línea negra punteada es $y = x$

Análisis de factores

ANÁLISIS DE FACTORES

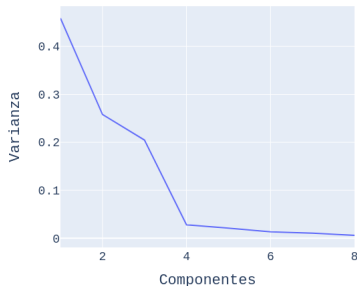
- Se utilizó la matriz de correlación **R**
- Calculando los valores propios obtenemos

eigenvalues = (3.669, 1.637, 2.064, 0.224, 0.169, 0.048, 0.083, 0.106)

Siguiendo el criterio de Kaiser tomamos $m = 3$.

ANÁLISIS DE FACTORES

Gráfico de codo



Varianza acumulada

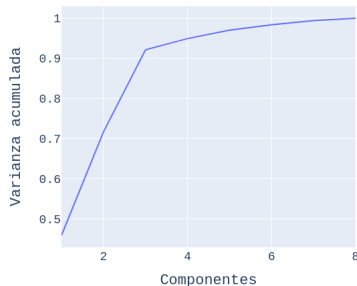


Figura: Gráfico de codo de las primeras componentes del análisis de factores. se puede observar que $m = 3$ factores superan el 90 % de varianza acumulada en los datos.

Figura: Gráfica de varianza acumulada. Un gran porcentaje de la varianza es explicada por las primeras 3 componentes principales.

ANÁLISIS DE FACTORES

- Aplicando el método por componentes principales, determinamos la matriz de cargas L .
- En amarillo se muestran aquellos valores con mayor aporte a cada factor.
- Los factores sin rotación no tienen una interpretación sencilla.

ANÁLISIS DE FACTORES

Solución por factor principal:

- Aplicando una rotación hasta obtener un error de $\epsilon = 0.01$, se obtiene una matriz de cargas interpretable.
 - El primer factor corresponde a todos los indicadores de desempeño en la cancha del jugador.
 - El segundo factor están las dos variables que describen la talla de los jugadores.
 - El tercer factor describe la experiencia de los jugadores.

ANÁLISIS DE FACTORES

Variable	F1	F2	F3	Comunalidades h_i^2	Varianzas Específicas ψ
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: 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.

POSICIÓN DE JUEGO

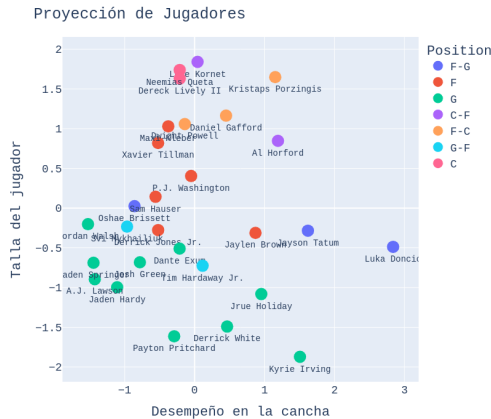


Figura: *factor scores* obtenidos con mínimos cuadrados ordinarios $\hat{\mathbf{f}}_j = (\mathbf{L}'_z\mathbf{L}_z)^{-1}\hat{\mathbf{L}}_z\mathbf{z}_j$

POR EQUIPOS

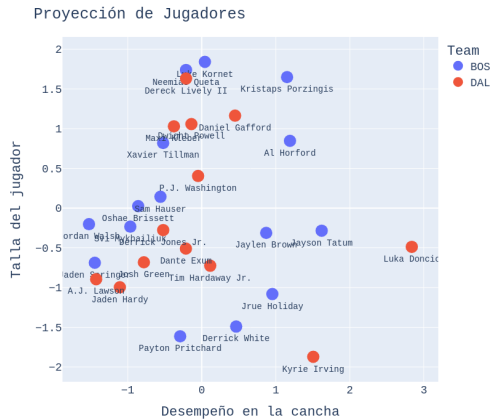


Figura: *factor scores* obtenidos con mínimos cuadrados ordinarios.

ANÁLISIS DE FACTORES

Solución por máxima verosimilitud: Para obtener factores más interpretables se aplicó una rotación *varimax*

Variable	F1	F2	F3	Comunalidades h_i^2	Varianzas Específicas ψ_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: Cargas determinadas por el método de máxima verosimilitud con una rotación *varimax*. 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 %

ANÁLISIS DE FACTORES

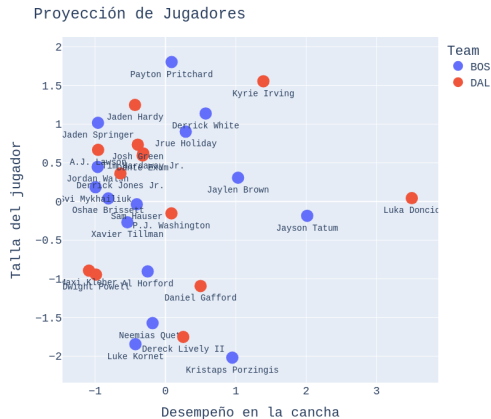


Figura: *factor scores* obtenidos con mínimos cuadrados ponderados $(\hat{f}_j = \hat{L}'_z \hat{\Psi}^{-1} \hat{L}_z)^{-1} \hat{L}'_z \hat{\Psi}^{-1} z_j$

ANÁLISIS DE FACTORES

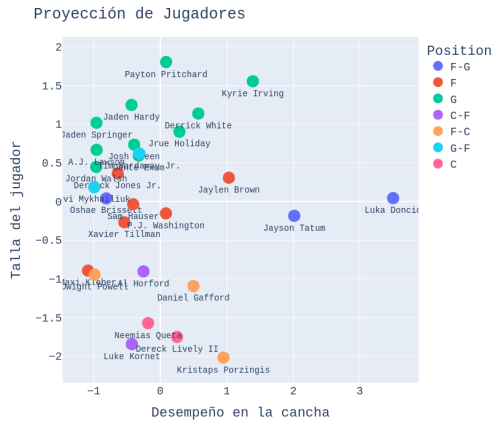


Figura: *factor scores* obtenidos con mínimos cuadrados ponderados

RESULTADOS DE LAS FINALES

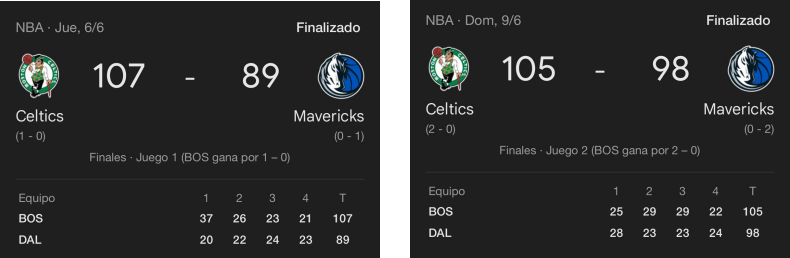


Figura: Resultados del marcador del primer y segundo juego de la final de Boston vs. Dallas en los PlayOffs 2023-2024

RESULTADOS DE LAS FINALES

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: Estadísticas de cada jugador en el primer juego de las finales de los PPlayoffs 2023-2024.

Conclusiones

CONCLUSIONES

Aspectos técnicos

- Los resultados describen los datos mediante tres variables latentes de manera exitosa con ambos métodos de rotación.
- El porcentaje de varianza original recuperado por estos factores supera el 80 %, indicando una buena aproximación de la matriz de covarianzas.
- No se rechaza la hipótesis de que los datos provienen de una distribución normal multivariada.
- El análisis de factores por el método de factor principal muestra que los factores se cargan en variables con alta correlación entre ellas, destacando características esperadas de los jugadores.
- El método de Máxima verosimilitud con rotación varimax produce una matriz de cargas igualmente interpretable y describe los mismos rasgos que el método de factor principal.





CONCLUSIONES

En relación a la final

- Las proyecciones en baja dimensión revelan un desbalance considerable entre los equipos: Boston tiene una ventaja en número y diversidad de jugadores con buenas estadísticas, mientras que Dallas solo cuenta con dos jugadores destacables.
- La distribución de los datos por posiciones muestra cómo la talla de cada jugador influye en sus posiciones de juego, sugiriendo posibles estrategias para introducir nuevos jugadores.
- Los datos fueron recabados antes de los primeros dos juegos de las finales, en los cuales Boston ganó ambos juegos, reflejando el desbalance en el desempeño de los jugadores en la cancha.

Referencias

REFERENCIAS I

-  Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. Prentice Hall.
-  <https://www.sportingnews.com/mx/nba?gr=www>
-  <https://www.britannica.com/biography/Karl-Pearson>
-  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

¡Gracias!