

# **Evaluación de la influencia de la llegada de LeBron James a Los Angeles Lakers: Análisis comparativo por temporada.**

IECD421: Visualización de datos

Bastián Barraza

---

## **1. Introducción**

LeBron James es uno de los atletas más dominantes de su generación y el año 2018 ha elegido dejar a los Cleveland Cavaliers para unirse a Los Angeles Lakers. Su cambio de equipo generó ventas de entradas a un precio 4 veces mayores en los tickets del Staples Center. De hecho, según el artículo de Golden y Chami en la revista CNBC (2018) [1], los precios del estadio de Los Ángeles aumentaron un 427 %. Mas aún Steinberg (2018) [2] asegura que , durante su primer partido en pretemporada, ESPN tuvo 1.17 millones de visitas, lo cual es la tercera audiencia más grande en un partido de pretemporada desde el año 2009.

De este modo, la problemática que se aborda en el presente informe es comprender si la llegada de LeBron James a los Los Angeles Lakers realmente impacta en la producción de puntos del equipo. Esta cuestión adquiere especial relevancia considerando el éxito que culminó en el campeonato obtenido por los Lakers en 2020. ¿Fue la presencia de LeBron un factor determinante en este logro, o existen otros factores en juego?. Dado esto, el objetivo planteado es evaluar si la llegada de LeBron James a los Los Angeles Lakers impacta significativamente en la producción de puntos del equipo, a través de la comparación de las temporadas previas y posteriores a su incorporación.

Para cumplir con los objetivos de la investigación, se emplearon análisis de series temporales y análisis de varianza (ANOVA) con el fin de evaluar patrones temporales y diferencias en el rendimiento del equipo en términos de puntos. El análisis temporal se llevó a cabo para examinar la producción promedio de puntos del equipo en intervalos de 10 segundos antes y después de la incorporación de LeBron James. Además, se utilizó el análisis de varianza para identificar diferencias significativas en la producción de puntos en los partidos de las temporadas de los Los Angeles Lakers.

Para llevar a cabo estos análisis, se hizo uso del entorno de programación estadística R en combinación con las librerías ggplot2, gridExtra y dplyr para el procesamiento y visualización de los datos. Asimismo, las librerías befast y agricolae fueron empleadas para realizar el análisis de series temporales y ANOVA, respectivamente.

Esta investigación se basa en la hipótesis de que existe una diferencia significativa en la producción de puntos de los Los Angeles Lakers antes de la llegada de LeBron James y después de su incorporación al equipo.

## **2. Objetivos**

### **2.1. Objetivo general**

Evaluar la influencia de LeBron James en el rendimiento de los Los Angeles Lakers en términos de la cantidad de puntos anotados por el equipo durante las temporadas 2017-18, 2018-19 y 2019-20.

## 2.2. Objetivos específicos

- Analizar la cantidad promedio de puntos anotados por el equipo en intervalos de 10 segundos, antes de la llegada de LeBron James y durante los dos años posteriores a su incorporación al equipo.
- Determinar si existen diferencias significativas en la cantidad de puntos anotados en cada partido de la temporada en Los Angeles Lakers.

## 3. Metodología

Para lograr los objetivos de este informe, se plantea analizar la serie temporal de la cantidad de puntos medios anotados en cada partido en intervalos de 10 segundos. Además, con el fin de determinar si la cantidad total de puntos por partido difiere significativamente entre las distintas temporadas, se emplea un análisis de varianza (ANOVA). Cada uno de estos análisis se lleva a cabo en el entorno de RStudio, versión 4.3.1, y las visualizaciones se crean utilizando la biblioteca ggplot2, versión 3.4.4.

### 3.1. Análisis temporal

La serie correspondiente a la media de cantidad de puntos por partido, realizada en intervalos de 10 segundos, se analiza mediante la librería 'bfast' en su versión 1.6.1. Para verificar la existencia de diferencias en cada temporada, se lleva a cabo un análisis comparativo con las series de cada temporada (2017-18, 2018-19 y 2019-20).

Además, se utilizan las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) para analizar la estructura de autocorrelación en la serie. Para descomponer la serie en sus componentes fundamentales, como la tendencia ( $T_t$ ), la estacionalidad ( $S_t$ ) y el error ( $e_t$ ), se realiza una descomposición de la serie. Por último, se emplean puntos de quiebre (breakpoints) para identificar momentos en los que la cantidad de puntos experimenta cambios significativos en su comportamiento.

### 3.2. Análisis ANOVA

Para obtener conclusiones más sólidas, se emplea un análisis de varianza (ANOVA) para evaluar si la cantidad de puntos por partido en la temporada 2017-18 (anterior a la llegada de LeBron James) difiere significativamente de las temporadas 2018-19 y 2019-20. Para llevar a cabo este análisis, se utiliza la biblioteca 'agricolae' en su versión 1.3.7.

La fórmula del análisis de varianza está dada por:

$$X_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

Donde:

- $X_{ij}$  es la cantidad de puntos en el partido  $i$  del año  $j$ .
- $\mu$  es la media global de todos los partidos jugados.
- $\alpha_i$  es el efecto del año  $i$ .
- $\epsilon_{ij}$  es el error aleatorio asociado con la observación  $X_{ij}$ .

Una vez completado el test ANOVA, se procede a verificar los supuestos de normalidad y homogeneidad de la varianza mediante las pruebas de Shapiro-Wilk y Levene, respectivamente. Se considera que los supuestos se cumplen si el valor de  $p$  es mayor que  $\alpha = 0,05$ . Luego, en caso de encontrar una diferencia significativa en el test ANOVA, se utiliza la prueba de Tukey para identificar las diferencias significativas entre pares de temporadas.

## 4. Resultados

En esta sección se visualizan los resultados obtenidos por el análisis temporal en cada temporada del equipo y de la prueba ANOVA.

### 4.1. Análisis temporal

Primero, en la figura 1, se muestra la serie temporal de la media de puntos en cada partido de la temporada, con intervalos de 10 segundos, categorizados por año de la temporada. Observamos que existe una tendencia al aumento en la cantidad de puntos por partido a medida que avanza el tiempo del partido. Además, cabe destacar que al comienzo de cada cuarto, la cantidad media de puntos tiende a aumentar. Se observan cambios significativos de tendencia en la temporada 2019, donde la cantidad de puntos aumenta notablemente al comenzar el segundo y último cuarto.

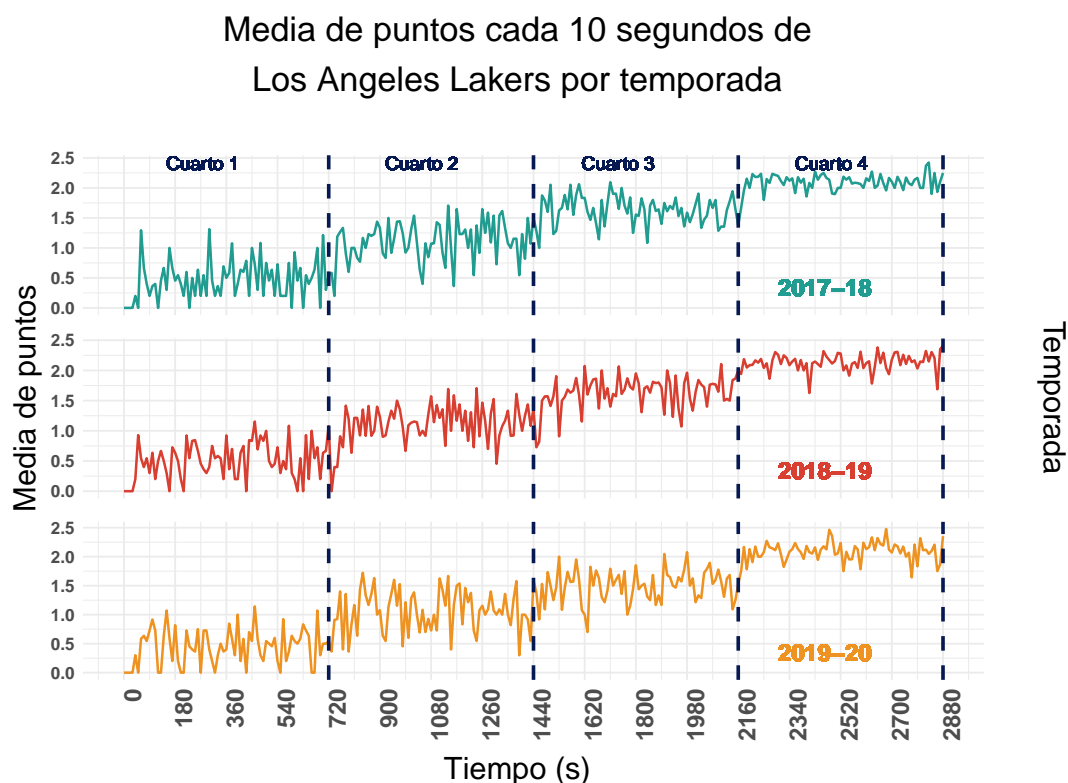


Figura 1: Serie temporal de la media de puntos con intervalos de 10 segundos en cada temporada.

Por otro lado, al analizar los gráficos de las funciones de autocorrelación (ACF), se observa una correlación positiva en todos los lags de tiempo para cada temporada, ya que los valores de autocorrelación se encuentran por encima del intervalo de confianza  $(-0.05, 0.05)$ . En general, la autocorrelación parcial (PACF) indica que no existe una influencia directa entre las observaciones en los lags superiores a 1 y la observación actual. No obstante, se aprecia influencia en lags menores a 1, y este patrón se repite en todas las temporadas. Estos resultados se pueden observar en la figura ??.

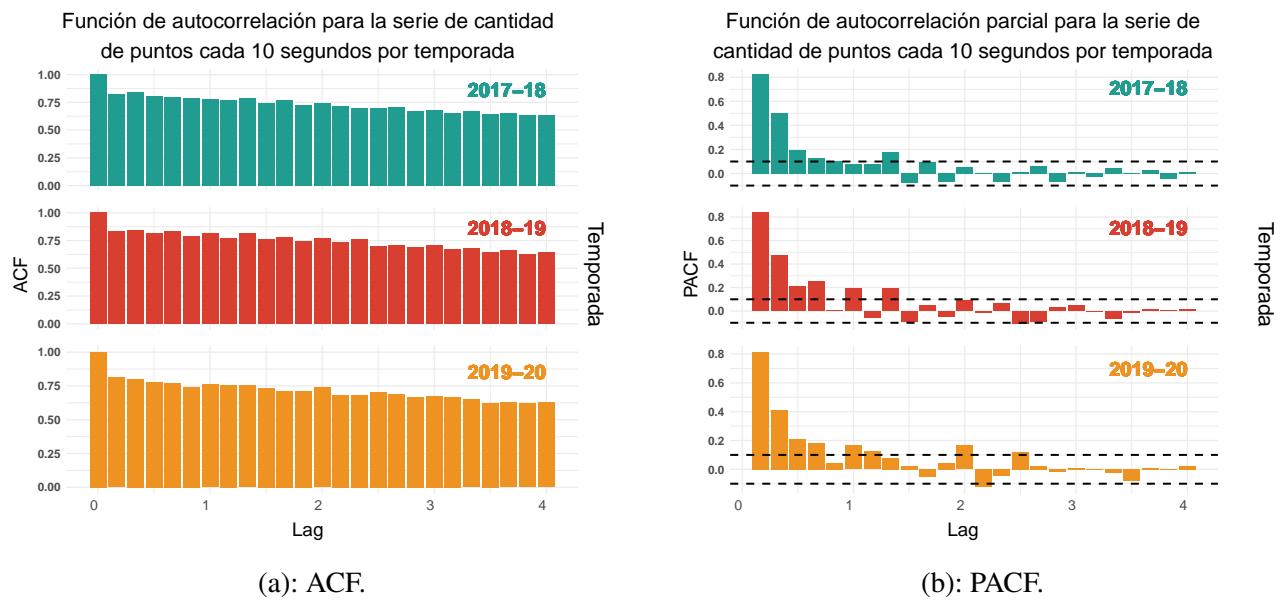


Figura 2: Funcion de autocorrelación y autocorrelación parcial para la cantidad de puntos realizados en intervalos de 10 segundos.

Las figuras 3, 4 y 5 muestran la descomposición de la serie de la media de puntos en las temporadas 2017-18, 2018-19 y 2019-20 en sus componentes fundamentales. Es relevante destacar que en las tres figuras se aprecia una estacionalidad en la cantidad de puntos. Además, se observan cambios significativos cerca del final o al inicio de cada cuarto, lo que sugiere la capacidad del equipo para anotar más puntos en las proximidades de cada descanso o después de este. No se identifican diferencias en el aumento de la cantidad media de puntos anotados en cada temporada después de estos cambios.

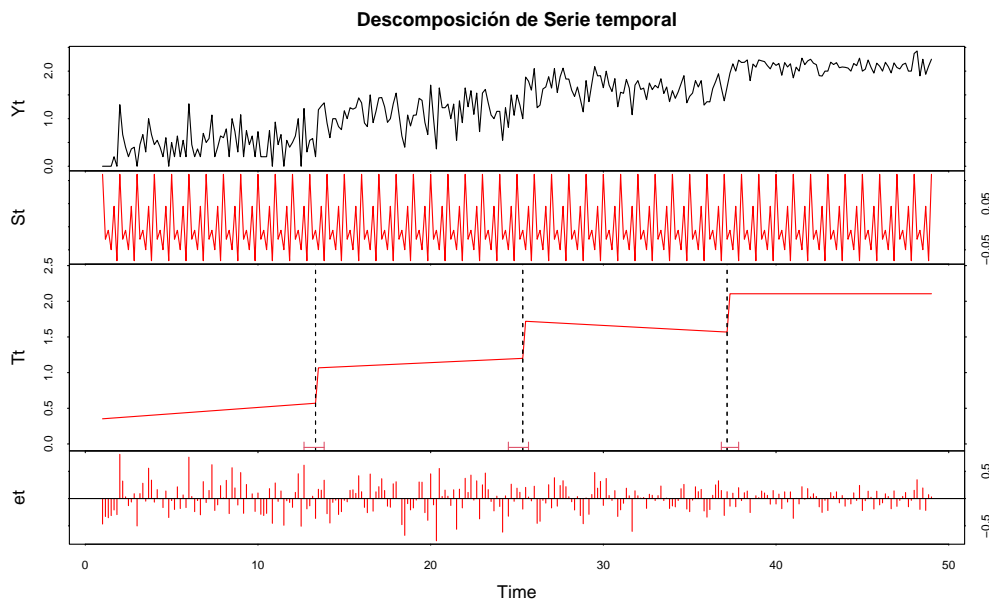


Figura 3: Descomposición de la serie temporal en la temporada 2017-18.

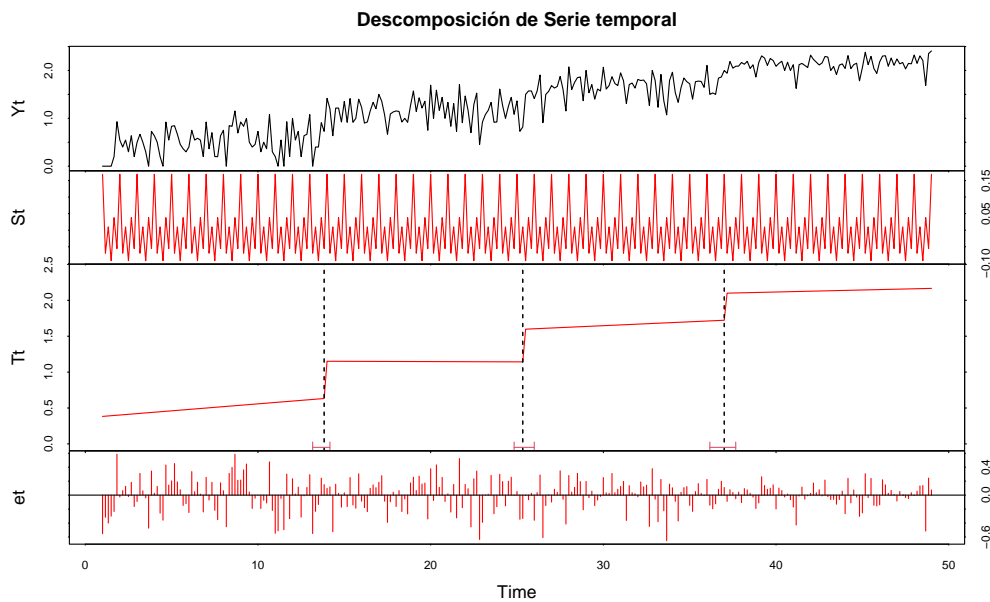


Figura 4: Descomposición de la serie temporal en la temporada 2018-19.

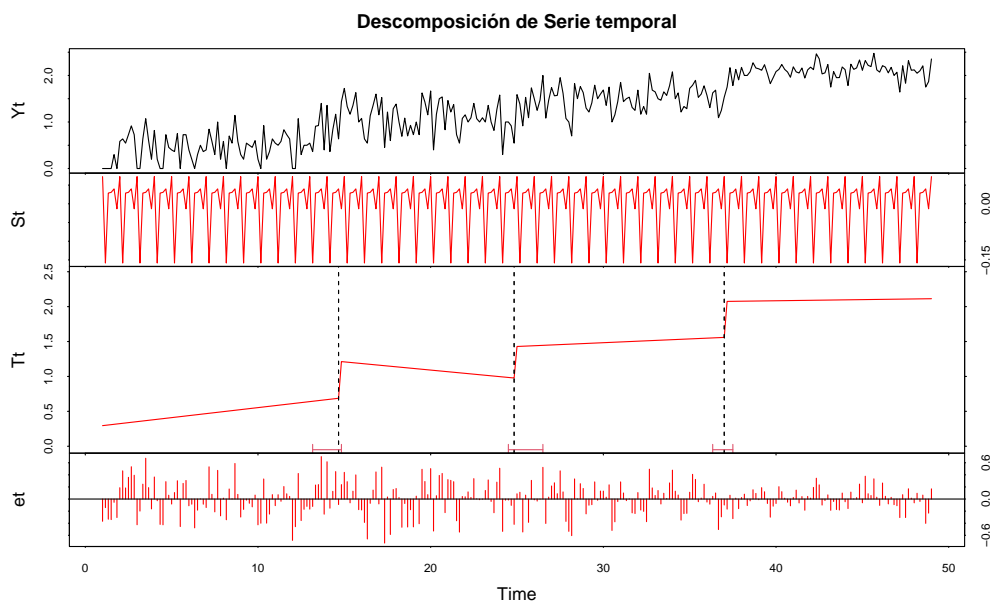


Figura 5: Descomposición de la serie temporal en la temporada 2019-20.

Los quiebres notables que se observan en los gráficos anteriores de descomposición de la serie están detallados en las tres tablas de la figura 1, junto con sus intervalos de confianza. El final de cada cuarto se produce en los momentos 72, 144, 216 y 288. Es relevante destacar que el final del segundo y tercer cuarto coincide con uno de los puntos de quiebre en cada temporada, sin embargo, el final del primer cuarto solo se alinea con el punto de quiebre en la temporada 2017-18, mientras que en las temporadas 2018-19 y 2019-20, los quiebres se producen un poco más tarde. Esto sugiere que los cambios en la cantidad de puntos ocurren principalmente en la primera parte del segundo cuarto y al final del segundo y tercer cuarto o al comienzo del tercer y cuarto cuarto.

2.5 %	Breakpoints	97.5 %	2.5 %	Breakpoints	97.5 %	2.5 %	Breakpoints	97.5 %
71	75	78	74	78	80	74	83	84
142	147	149	144	147	151	142	144	154
216	218	222	212	217	221	213	217	220

(a): Temporada 2017-18

(b): Temporada 2018-19

(c): Temporada 2019-20

Cuadro 1: Puntos de cambio de tendencia y sus intervalos de confianza para la serie temporal.

## 4.2. Análisis ANOVA

Para analizar las diferencias en las medias de cada temporada, se lleva a cabo un análisis ANOVA. En la figura 6, se presentan los boxplots que muestran el total de puntos anotados por los Angeles Lakers en cada temporada. Se puede observar que la mediana es ligeramente mayor en los años 2018 y 2019, y además, el total de puntos en estos años alcanza máximos más significativos que en la temporada 2017.

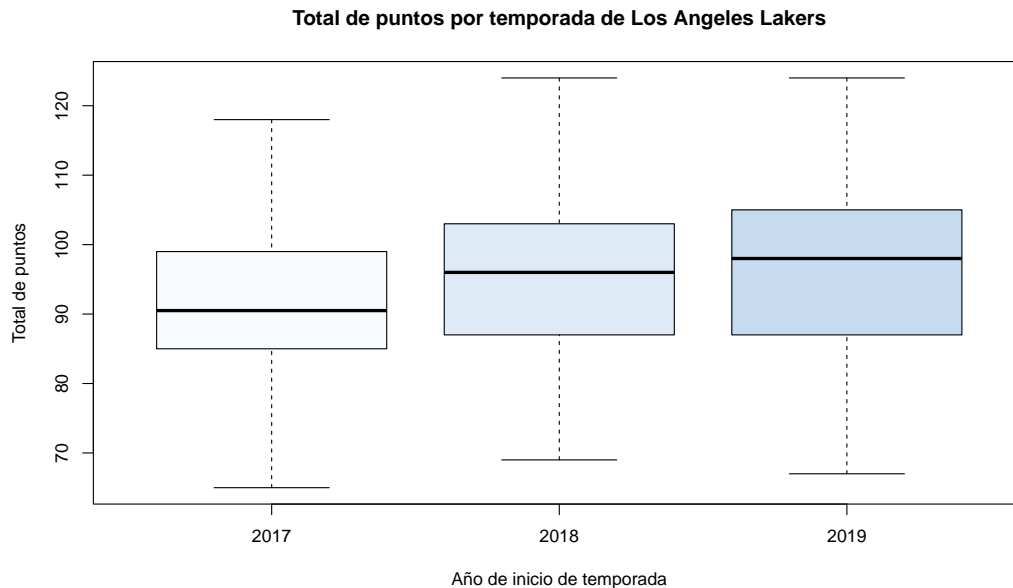


Figura 6: Boxplot del total de puntos anotados en cada temporada del equipo.

Asimismo, en la tabla 2, se muestran la cantidad de puntos y la desviación estándar de la cantidad de puntos por partido en cada temporada. A partir de estos datos, se puede observar que hubo una mayor cantidad de puntos anotados en los partidos de las temporadas 2018 y 2019. Además, se destaca que la mayor variación en los datos se encuentra en la temporada 2019.

Temporada	Media de puntos	Desv. Estándar
2017	91.439	11.997
2018	95.475	11.396
2019	95.690	12.354

Cuadro 2: Estadísticas descriptivas por temporada de la cantidad de puntos realizados por partido Los Angeles Lakers.

Después de llevar a cabo el test ANOVA, observamos que, estrictamente hablando, sí existen diferencias significativas entre las medias de los grupos al establecer un nivel de significancia de  $\alpha = 0,05$ , ya que el valor p es de 0.04. Sin embargo, es relevante tener en cuenta que el valor p se encuentra muy cercano al nivel de significancia, lo que sugiere que el grado de diferencia entre las medias no es muy pronunciado. El resultado del test se presenta en la tabla 3

Variable	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Year	2	915	457.5	3.23	0.0413*
Residuals	232	32864	141.7		

Cuadro 3: Resultados del test de análisis de varianza (ANOVA).

#### 4.2.1. Supuestos

Para verificar los supuestos de normalidad y homogeneidad de varianza, se emplearon los tests de Shapiro-Wilk y Levene, respectivamente. Los resultados presentados en la tabla 4 sugieren que se cumplen los supuestos necesarios para que el modelo ANOVA funcione correctamente. Esta comprobación se refuerza al observar el boxplot de los residuos del modelo en la figura 7, donde se aprecia que el 50 % de los datos se encuentra en el centro de la distribución de los residuos y sin detectarse valores atípicos o asimetrías significativas en la figura.

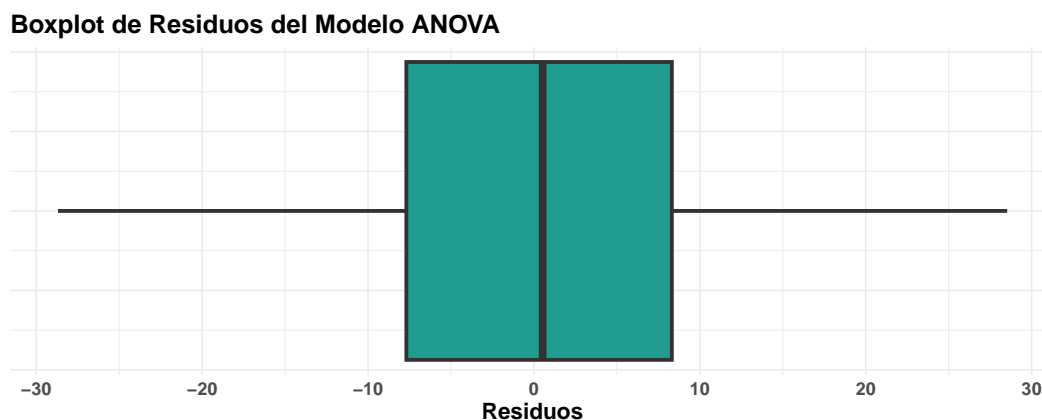


Figura 7: Boxplot de residuos del modelo de análisis de varianza (ANOVA).

Prueba	Estadístico	P-value
Shapiro-Wilks	0.994	0.632
Levene	0.783	0.458

Cuadro 4: Resultados de pruebas de normalidad y homocedasticidad.

#### 4.2.2. Post-Anova

Dado que se han encontrado diferencias significativas en el test, se ha planteado la realización de la prueba de Tukey para identificar diferencias significativas entre pares de temporadas. Los resultados se detallan en la tabla 5. Es importante destacar que la diferencia en la cantidad de puntos anotados entre los años 2017 y 2018, así como la diferencia entre 2017 y 2019, presenta valores de p por encima de  $\alpha = 0,05$ , lo que, estrictamente hablando, indica que no existen diferencias significativas entre estos pares de años. No obstante, dado que estos valores se encuentran cerca del nivel de significancia, se puede inferir que, aunque no existan diferencias significativas, la

cantidad de puntos entre estos pares de años no es muy similar. Por otro lado, el par de temporadas 2018 y 2019 no muestra una diferencia significativa en el total de puntos por partido.

Año	Diferencia	P-Value
2017 - 2018	-4.037	0.079
2017 - 2019	-4.251	0.073
2018 - 2019	-0.215	0.993

Cuadro 5: Resultados de la prueba de Tukey en Post-Anova.

## 5. Conclusión

Se ha llevado a cabo un análisis de la serie temporal de la media de puntos en los partidos de las temporadas 2017-18, 2018-19 y 2019-20. Los objetivos se han cumplido, ya que se logró evaluar la influencia de LeBron James en su llegada a Los Angeles Lakers mediante el análisis de series temporales y pruebas ANOVA.

Sin embargo, la hipótesis planteada se corrobora parcialmente, ya que, aunque existen diferencias significativas en la cantidad de puntos anotados en cada temporada, no se logran identificar diferencias significativas entre las temporadas en pares.

Se observó una tendencia al aumento de puntos a lo largo de los partidos, con cambios notables cerca del final o inicio de cada cuarto. Además los quiebres identificados en las tendencias de las series temporales, cerca del final o inicio de cada cuarto, proporcionan información sobre momentos clave en los que el equipo tiende a anotar más puntos.

El análisis ANOVA han revelado diferencias significativas entre las medias de las temporadas, aunque estas diferencias son relativamente pequeñas y se acercan al nivel de significancia. Ahora bien, la prueba de Tukey ha identificado que las diferencias entre las temporadas 2017-18 y 2018-19, así como entre 2017-18 y 2019-20, no son estadísticamente significativas, pero podrían indicar que la cantidad de puntos entre estos pares de años no es muy similar.

Como nota final, se puede inferir que si bien, la llegada de LeBron James al equipo mostró un aumento en la producción de puntos, este incremento no alcanzó un nivel estadísticamente significativo.

## Referencias

- [1] Eric Golden Jessica; Chemi. "The LeBron James Effect Is Boosting Lakers Ticket and Merchandise Sales". En: *CNBC* (oct. de 2018). URL: <https://www.cnbc.com/2018/10/07/the-lebron-james-effect-is-boosting-lakers-ticket-and-merchandise-sales.html>.
- [2] Leigh Steinberg. "The 'LeBron James Effect': How Will It Impact The Lakers?" En: *Forbes* (2018). URL: <https://www.forbes.com/sites/leighsteinberg/2018/10/31/the-lebron-james-effect/?sh=6f64b0014657>.

-