Proyecto Métodos no Paramétricos:

Laura Silvana Alvarez[1], Daniel Eduardo Castro[2], Alejandro Coronado Sarmiento[3], Luisa Fernanda Guantiva Vargas[4], Laura Johana Lizarazo[5]

21 de agosto de 2024

Resumen: El baloncesto es uno de los deportes más practicados del mundo, en la actualidad cuenta con una gran difusión. Se juegan numerosas ligas y campeonatos en el mundo entero y en Estados Unidos, se disputa la NBA, considerada la competición más prestigiosa en el mundo. En los últimos años la liga estadounidense ha aumentado sus ingresos y por lo tanto los salarios de los jugadores por lo cual, por medio de métodos estadísticos no paramétricos analizaremos los factores que afectan los salarios de los jugadores y sus diferencias por temporadas.

Palabras clave: Métodos, distribución, no parámetrico, test, factores, correlación, hipótesis, pareadas.

1. Introducción

La **National Basketball Association** (mejor conocida como NBA) es la liga profesional de baloncesto más reconocida a nivel mundial. Esta asociación fue fundada en 1946, año en el cual se realizo una liga con 11 equipos profesionales. A medida que los años fueron avanzando, algunos equipos desaparecieron y otros equipos se fueron uniendo a la NBA. Hoy en día un total de 30 equipos hacen parte de la liga.

Estos 30 equipos se dividen en dos conferencias (este y oeste), donde cada conferencia tiene una misma cantidad de equipos. Durante la temporada regular cada equipo compite en 82 juegos de 48 minutos, y a diferencia de otros deportes solo existen dos posibles resultados de un partido: victoria o derrota. En caso de empatar en el tiempo regular se juega un tiempo suplementario de 5 minutos para decidir el ganador del partido. Al final de la temporada regular los 8 equipos de cada conferencia con mejor porcentaje de victoria acceden a la siguiente ronda, conocida como los *Playoffs*.

En los *playoffs* se emparejan los equipos de cada conferencia en eliminatorias directas. Estas eliminatorias son a 7 partidos, es decir el primer equipo en ganar 4 partidos accede a la siguiente ronda. Las siguientes rondas tienen el mismo formato de eliminatorias a 7 partidos, y así al final solo restan dos equipos (que serian los campeones de cada conferencia), los cuales se enfrentan también en una eliminatoria a 7 partidos para definir el campeón.

Como en cualquier otro deporte los equipos que tienen mejor desempeño son los equipos que tienen las plantillas con los mejores jugadores. Un equipo es exitoso usualmente cuando tiene una plantilla llena de jugadores de alto nivel y experiencia, o cuando tiene dentro de sus plantilla uno o dos jugadores *All-Star*. Los jugadores *All-Star* son elegidos cada año por los aficionados, y son lo que tienen un desempeño exageradamente alto en la liga. Los *All-Star* también son los jugadores más reconocidos a nivel mediático, como por ejemplo lo fueron grandes individuales como *Michael Jordan, Kobe Bryant*, o *Shaquille O'Neal*. Hoy en dia los *All-Star* mas reconocidos son *Stephen Curry, LeBron James, Giannis Antetokounmpo*, y *James Harden*.

Como es de esperarse estos jugadores *All-Star* exigen un salario más alto para hacer parte de un equipo de la NBA. Esto generó una problemática en la NBA ya que los equipos con más dinero podían tener plantillas mucho más fuertes a nivel deportivo que el resto de equipos. Por tal razón la NBA instauró en 1983 un tope salarial, el cual se calcula todas las temporadas a partir de los ingresos generados por la NBA en materia de básquet. El tope salarial es un límite impuesto por la liga al gasto de los equipos en materia de sueldos, el cual es igual para todos los equipos. Esto busca garantizar cierta competitividad en la NBA y también mejorar la estructura financiera de la liga. Cuanto más dinero se genere, más tienen los equipos para gastar en sueldos, lo que genera una estructura bastante sustentable en materia de finanzas. Para la temporada 2017–2018 el número límite fue de 99.093.000 de dólares por equipo.

[1]lsalvarezlu@unal.edu.co
[2]decastrome@unal.edu.co
[3]deacoronadosa@unal.edu.co
[4]lfguantivav@unal.edu.co
[5]ljlizarazore@unal.edu.co

Sin embargo las diferencias salariales siguen existiendo entre los equipos, y estas diferencias son consecuencia de varios factores característicos de cada jugador. En este proyecto se busca estudiar que factores afectan más los salarios de los jugadores, aplicando métodos estadísticos no paramétricos en una base datos que contiene los salarios de cada jugador de la NBA considerando las temporadas desde 2011-2012 hasta 2018-2019.

2. Marco Teórico

Algunos de los métodos aplicados para hacer el análisis de los datos se presentan a continuación:

2.0.1. Normalidad

- **Kolmogorov-Smirnov**
  Sea $X_1, ..., X_n$ una muestra aleatoria de $X \sim F(x)$ y $X_{(1)}, ..., X_{(n)}$ las estadísticas de orden. Se plantea la hipótesis:

$$H_0 : F(x) = F_0(x) \quad vs. \quad H_a : F(x) \neq F_0(x)$$

  Donde $F_0(x)$ es una distribución de interés, en este caso la distribución normal. Además se define la variable auxiliar:

$$S_n(x) = \begin{cases} 0 & x < X_{(1)} \\ k/n & X_{(k)} \leq X < X_{(k+1)} \\ 1 & x \geq X_{(n)} \end{cases}$$

  A partir de lo anterior se define la estadística de prueba

$$\begin{aligned} D &= máx_{i=1,\ldots,n}|S_n(x) - F_0(x)| \\ &= máx_{i=1,\ldots,n}\left|\frac{i}{n} - F_0(X_{(i)})\right| \end{aligned}$$

  Para la toma de decisión se rechaza H_0 si $D > d_\alpha$.

- **Lilliefors**
  Este test se basa en una modificación de la estadística de Kolmogorov-Smirnov. En este caso se definen:

$$D^+ = máx_{i=1,\ldots,n}\left|\frac{i}{n} - F_0(X_{(i)})\right|$$

$$D^- = máx_{i=1,\ldots,n}\left|\frac{i-1}{n} - F_0(X_{(i)})\right|$$

  Y la estadística D con la que se toma la decisión es:

$$D = máx(D^+, D^-)$$

- **Jarque Bera**
  El test de Jarque Bera es un estadístico que involucra la curtosis y la asimetría. Está definido como:

$$JB = n \cdot \left[\frac{s^2}{6} + \frac{(k-3)^2}{24}\right] \sim \chi^2_2$$

  Donde n es el tamaño de muestra, s el coeficiente de asimetría y k la curtosis.
  Se rechaza la hipótesis nula de normalidad si $JB_c < \chi^2_{(2,\alpha)}$ con un nivel de significancia de α.

2.0.2. Test del Signo para muestras pareadas

Sean $(X_1,\, X_2)$ una variable aleatoria bivariada y $Y = X_1 - X_2$ donde Y tiene mediana θ, es decir, es un valor tal que $F_Y(\theta) = \frac{1}{2}$ (Si Y es continua) o $P(Y \leq \theta) \geq \frac{1}{2}$ y $P(Y \geq \theta) \geq \frac{1}{2}$ (Si Y es discreta).

La estadística de prueba y región de rechazo dada para Y son:

$$S = \sum_{i=1}^{n} \varphi_i \quad \text{donde} \quad \varphi_i = \begin{cases} 1 \text{ si } y_i > \theta_0 \\ 0 \text{ si } y_i < \theta_0 \end{cases}$$

Bajo la hipótesis $H_0 : \theta = 0$, se tiene $P(\varphi_I = 1) = P(Y_i > 0) = \frac{1}{2}$ entonces:

$$S \sim Bin\left(n, \frac{1}{2}\right)$$

La región de rechazo para esta hipótesis esta dada por:

+ Para $H_a : \theta > 0$, se rechaza H_0 si $S \geq k$, donde k es $P(S \geq k) = \alpha$ con α el nivel de significancia.
+ Para $H_a : \theta < 0$, se rechaza H_0 si $S \leq k$, donde k es $P(S \leq k) = \alpha$ con α el nivel de significancia.
+ Para $H_a : \theta \neq 0$, se rechaza H_0 si $S \leq k_1$, tal que k_1 es $P(S \leq k_1) = \alpha/2$ o si $S \geq k_2$, tal que k_2 es $P(S \geq k_2) = \alpha/2$, con α el nivel de significancia.

2.0.3. Test del Kruskal-Wallis

El test de Kruskal-Wallis es un test no paramétrico que extiende la idea del test de Mann-Whitney, permitiendo la comparación de $k \geq 2$ grupos de datos independientes, se toman k muestras independientes de distribuciones continuas con mediana θ. La hipótesis es que la mediana de las k distribuciones es igual, es decir:

$$\begin{aligned} &H_0 : \theta_1 = \theta_2 = \cdots = \theta_k \\ &H_a : \theta_i \neq \theta_j \quad \text{, Para algún } i \neq j \end{aligned}$$

La estadística de prueba esta dada por:

$$H = \frac{12}{N(N+1)} \left(\sum_{j=1}^{k} \frac{R_j^2}{n_j} \right) - 3(N+1)$$

Donde:

- $N = \sum_{j=1}^{k} n_j$
- R_j es la suma de los rangos de la j-esimo grupo.

La región de rechazo es:

+ Si $k \leq 5$ y los n_j son pequeños, se usa la distribución exacta.
+ Si las muestras son grandes, bajo H_0, $H \sim \chi^2_{(k-1)}$ se rechaza H_0 al nivel α si $H > \chi^2_{(1-\alpha,k-1)}$

Cuando hay presencia de empates a cada observación correspondiente se le asigna el promedio de los rangos iniciales. Adicionalmente, se realiza el ajuste de la estadística H dividiendo por la siguiente expresión:

$$C = \frac{\sum_{i=1}^{m} t_i^3 - t_i}{N^3 - N}$$

donde, t_i es el numero de empates en el i-esimo grupo y m es el numero de grupos con rangos empatados. Así la estadística corregida esta dada por:

$$H_{corr} = \frac{H}{1-C}$$

2.0.4. Test de Dunn

El $test\ de\ Dunn$ es el análogo no paramétrico de la prueba t para comparar medias. El test de Dunn es una manera efectiva de realizar inferencias simultáneas en parejas. Primero se combinan los datos, se jerarquizan y se hallan los rangos de las medias grupales. Las hipótesis que se manejan en esta prueba son:

H_0 : No existe diferencia entre los grupos.
H_a : Existe diferencia entre los grupos.

Se rechaza la hipótesis nula H_0 si:

$$\left| \frac{R_j}{n_j} - \frac{R_i}{n_i} \right| > Z_{1-\frac{\alpha}{2}} \sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_j} + \frac{1}{n_i} \right)}$$

donde:

- R_j es suma de los rangos del grupo j y R_i es suma de los rangos del grupo i.
- n_j número de datos del grupo j y n_i número de datos del grupo i.
- N es el número total de datos evaluados.
- $i, j = 1, 2, \cdots, k$

Si existen empates, se realiza la comparación con:

$$\left| \frac{R_j}{n_j} - \frac{R_i}{n_i} \right| > Z_{1-\frac{\alpha}{2}} \sqrt{\left(\frac{N(N+1)}{12} - B \right) \left(\frac{1}{n_j} + \frac{1}{n_i} \right)}$$

Donde B es:

$$B = \frac{\sum_{i=1}^{m} (t_i^3 - t_i)}{12(N-1)}$$

Donde, t_i es el numero de empates en i-esimo grupo.

2.0.5. Correlación

Test de Spearman:
El test de Spearman detecta la relación monótona (creciente o decreciente) entre dos variables, las cuales, no necesitan normalidad bivariada y pueden ser discretas. También es resistente a observaciones atípicas porque la variable D se define en términos de los rangos de cada variable de la siguiente forma:

Individuo	X	Y	R(X)	R(Y)	D
1	X_1	Y_1	$R(X_1)$	$R(Y_1)$	D_1
2	X_2	Y_2	$R(X_2)$	$R(Y_2)$	D_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	X_n	Y_n	$R(X_n)$	$R(Y_n)$	D_n

Cuadro 1: Variables, rangos y diferencia de rangos

En este test se hace el coeficiente de correlación de Pearson para los rangos.

$$\begin{aligned} R_s &= \frac{Cov(R(X), R(Y))}{S_{R(X)} S_{R(Y)}} \\ R_s &= \frac{\sum^m \left(R(X_i) - \frac{n+1}{2} \right) \left(R(Y_i) - \frac{n+1}{2} \right)}{\sqrt{\sum^m \left(R(X_i) - \frac{n+1}{2} \right)^2} \sqrt{\sum^m \left(R(Y_i) - \frac{n+1}{2} \right)^2}} \end{aligned}$$

Si no se presentan empates, la estadística se puede simplificar en términos de la diferencia de los rangos (D_i)

$$R_s = 1 - \frac{6 \sum_i D_i^2}{n^3 - n}$$

Se tiene que $E(R_s) = 0$ y $V(R_s) = \frac{1}{n-1}$. Por lo tanto la aproximación asintótica está dada por:

$$Z = \frac{R_s}{\sqrt{\frac{1}{n-1}}} \sim N(0,1)$$

Se rechaza H_0 al nivel α si $R_s > Z_{1-\frac{\alpha}{2}}$ o $R_s < Z_{\frac{\alpha}{2}}$.

3. Análisis Descriptivo

Para realizar el análisis de este proyecto, se verificó que no existieran registros con valores faltantes en la variable respuesta, se eliminaron registros cuyo equipo era desconocido y se hizo una limpieza en general de la base de datos.

Generando así que quedáramos con un total de 4474 registros para trabajar, junto con 5 columnas en total (Jugador, Salario, Salario con inflación, Temporada y Equipo). Donde la variable respuesta será el salario con inflación basándonos en el salario de 2020.

- En la variable respuesta, se obtiene que el mínimo es de $6.051 (Donte Greene en Memphis durante 2012-2013), el máximo de $38.320.489 (Stephen Curry en Golden State Warriors durante 2018-2019) y una media de $4.989.140 por temporada.
- Hay un total de 1162 jugadores, es decir, que hubo varios jugadores que estuvieron en más de una temporada.
- Hay un total de 30 equipos.
- Están registradas 8 temporadas, las cuales son desde la temporada 2011-2012 hasta la temporada 2018-2019.

También se analizó la variable salario con inflación por cada temporada y por cada equipo.

Para esto, se decidió realizar un acercamiento gráfico.

- Relación entre temporada y salario. Se ve que los salarios bajos no tienen un gran cambio, pero, sí existe variación entre los salarios altos.

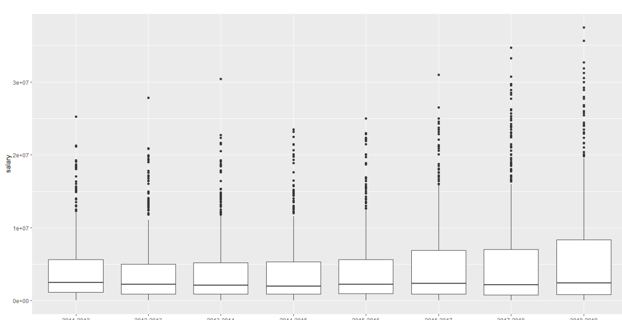

Figura 1: Relación salario y temporada

- Relación entre equipo y salario. Se ve diferencia de la variable respuesta entre los 30 equipos, dado que existen algunos con gran variabilidad dentro, mientras otros no tanto. Además, dentro de los equipos hay varios con observaciones anormales y extremas, por los jugadores con alto salario a comparación del resto del equipo.

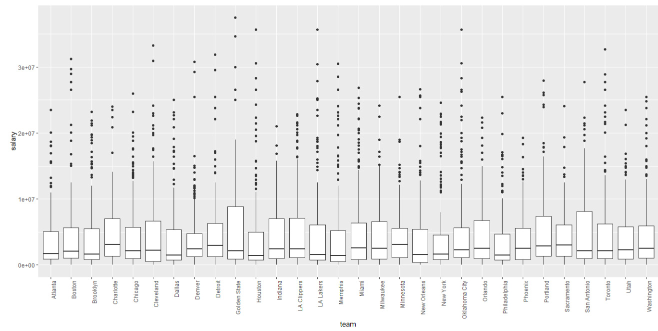

Figura 2: Relación salario y equipo

4. Aplicación

Tras hacer el análisis descriptivo, ahora se busca verificar el supuesto de normalidad de este conjunto de datos.

4.0.1. Normalidad

Para verificar si los datos distribuyen normal, se utilizaron algunas pruebas tales como Lilliefors. Primero se restó la media por temporada y equipo a cada uno de los salarios de los jugadores, dado que no todos los datos provenían del mismo lugar. Generando así que la variable a revisar sea X=Salario-Media.

```
E5 <- sqldf(
  "select team,
   avg(salary) as media, season
   from NBA
   group by team,season")
NBA2 <- sqldf(
  "select player,NBA.team,NBA.season,
     media, salary
     from NBA left join E5 on
     (NBA.team = E5.team) and (NBA.season
          = E5.season)")

x= NBA2$salary-NBA2$media
```

Antes de realizar las pruebas se hizo un acercamiento gráfico, y se calculó el sesgo y la curtosis del conjunto de datos.

```
skewness(x)
 1.740459
kurtosis(x)
 6.060951
```

```
#graficos
qqnorm(x)
hist(x)
```

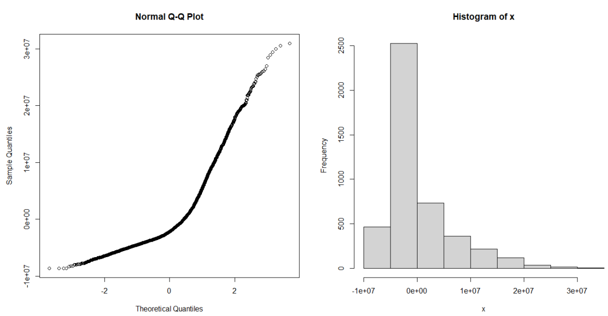

Figura 3: Gráficos de salarios

Con los resultados anteriores, puede que no haya normalidad en la variable X. Para revisar lo anterior se utilizan las pruebas de normalidad, tales como Lilliefors, Kolmogorov-Smirnov y Jarque Bera, con la hipótesis nula H_0 : Los datos provienen de una población con distribución normal contra H_a : Los datos no provienen de una población con distribución normal.

```
Lilliefors (Kolmogorov-Smirnov) normality
     test

data:  x
D = 0.17505, p-value < 2.2e-16
#
ks.test(x,"pnorm", mean(x),sd(x))

  One-sample Kolmogorov-Smirnov test

data:  x
D = 0.17505, p-value < 2.2e-16
alternative hypothesis: two-sided
#
jarque.bera.test(x)

  Jarque Bera Test

data:  x
X-squared = 4005.4, df = 2, p-value < 2.2
    e-16
```

Teniendo en cuenta los resultados presentados anteriormente, existe evidencia suficiente para afirmar que los datos no provienen de una población con distribución normal. A partir de este resultado se decide utilizar métodos no paramétricos para el análisis presentado a continuación.

4.0.2. Test del Signo para muestras pareadas

Los patrocinios son una parte importante del negocio de la NBA. Uno de los cambios más importantes en este aspecto, ocurrió en junio de 2015 cuando la NBA acabó con la larga asociación que tenían con *Adidas* y firmó un contrato multimillonario con *Nike*.

Para saber si este cambio tuvo un impacto positivo sobre el salario de los jugadores, se realiza el test del signo para muestras pareadas con el salario de los jugadores en la temporada 2014-2015 y en la temporada 2015-2016; cuando se hizo el cambio del contrato.

Primero se realiza un gráfico con las densidades de los salarios con cada contrato para comparar los salarios de los jugadores. Se cuenta con 403 jugadores que participaron en las dos temporadas (Figura 4).

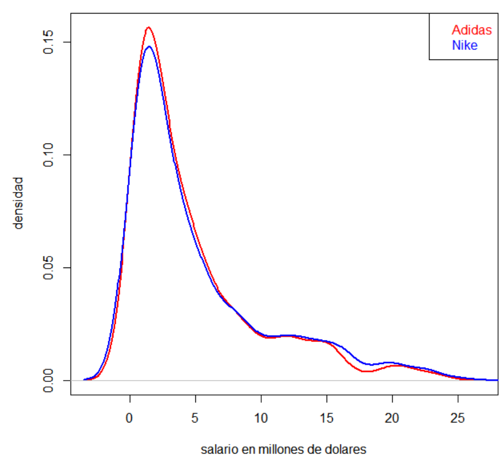

Figura 4: Comparación de patrocinadores

Se puede observar que en los salarios más bajos hay mayor densidad en el contrato con Adidas, mientras que en los salarios entre 15 y 20 millones dólares hay mayor densidad con el patrocinio de Nike. Con los demás salarios se observa una distribución similar con ambos patrocinadores.

Ahora, para comprobar si esta diferencia es significativa, se realiza el test planteando la hipótesis $H_0 : \theta_{2016} \leq \theta_{2015}$ vs. $H_a : \theta_{2016} > \theta_{2015}$.

```
SIGN.test(data2$salario2016, data2$
    salario2015, alternative = "greater",
      conf.level = 0.95)

# Dependent-samples Sign-Test
# data:  data2$salario2016 and  data2$
    salario2015
# S = 268, p-value < 2.2e-16
```

A partir de estos resultados hay evidencia para rechazar $H_0 : \theta_{2016} \leq \theta_{2015}$, por lo que el cambio de contrato significaría un aumento en el salario de los jugadores.

4.0.3. Test del Kruskal-Wallis

Usaremos el test de $Kruskall-Wallis$ para $k=8$ muestras independientes, que en este caso son temporadas de distribuciones continuas con mediana θ. La hipótesis que planteamos es que la mediana de las 8 temporadas son iguales, es decir:

$$H_0 : \theta_1 = \theta_2 = \cdots = \theta_8$$
$$H_a : \theta_i \neq \theta_j \quad \text{, Para algun } i \neq j$$

Con la ayuda del software **R** usando la función "$Kruskal.test$" obtenemos los siguientes resultados para nuestras 8 temporadas:

```
> kruskal.test(dsueldos ~ grupos, data =
    datos2)

  Kruskal-Wallis rank sum test

data:  dsueldos by grupos
Kruskal-Wallis chi-squared = 117.48, df =
    7, p-value < 2.2e-16
```

Después de realizar el test para comparar las medianas de los salarios de los jugadores en las 8 temporadas, con un nivel de confianza del $95\,\%$ podemos afirmar que existe evidencia estadística para rechazar la hipótesis nula H_0, es decir, existe diferencias entre las medianas de los salarios.

4.0.4. Test del Dunn

Basándonos en el resultado del test de $Kruskal-Wallis$, sabemos que existe alguna diferencia entre las medianas de los salarios en las 8 temporadas. Para ver más precisamente entre que temporadas existen estas diferencias realizaremos el $test\ de\ Dunn$ el cual nos permite realizar inferencias simultaneas sobre los 8 grupos. Las hipótesis que se evalúan son las siguientes:

$$H_0 : \theta_i = \theta_j \text{ Para algún } i \neq j$$
$$H_a : \theta_i \neq \theta_j \text{ Para algún } i \neq j$$

Usamos la función $dunn.test$ para evaluar las hipótesis y poder aceptar o rechazar por parejas la relación entre la mediana de los salarios.

```
dunn.test (dsueldos,g=grupos,method="
    bonferroni",
        kw=TRUE,label=TRUE,alpha=0.05)
```

En la siguiente tabla se presenta la comparación entre los grupos usando el $valor-P$ como indicador para evaluar las hipótesis:

```
+         kw=TRUE,label=TRUE,alpha=0.05)
  Kruskal-Wallis rank sum test

data: dsueldos and grupos
Kruskal-Wallis chi-squared = 117.483, df = 7, p-value = 0


                    Comparison of dsueldos by grupos
                              (Bonferroni)
Col Mean-|
Row Mean |          1          2          3          4          5          6          7
---------+----------------------------------------------------------------------------
       2 |  -1.642131
         |     1.0000
         |
       3 |  -2.304965  -0.662834
         |     0.2964     1.0000
         |
       4 |  -3.698364  -2.056233  -1.393399
         |    0.0030*     0.5566     1.0000
         |
       5 |  -5.914566  -4.272435  -3.609601  -2.216201
         |    0.0000*    0.0003*    0.0043*     0.3735
         |
       6 |  -7.929698  -6.287567  -5.624733  -4.231333  -2.015132
         |    0.0000*    0.0000*    0.0000*    0.0003*     0.6145
         |
       7 |  -7.537980  -5.895848  -5.233014  -3.839615  -1.623413   0.391718
         |    0.0000*    0.0000*    0.0000*    0.0017*     1.0000     1.0000
         |
       8 |  -5.917461  -4.275329  -3.612495  -2.219096  -0.002894   2.012237   1.620519
         |    0.0000*    0.0003*    0.0042*     0.3707     1.0000     0.6187     1.0000
alpha = 0.05
Reject Ho if p <= alpha/2
```

Figura 5: Tabla test de Dunn

Con estos resultados, podemos observar:

- Diferencias entre la temporada 1 y las temporadas 5, 6, 7 y 8.
- Diferencias entre la temporada 2 y las temporadas 5, 6, 7 y 8.
- Diferencias entre la temporada 3 y las temporadas 5, 6, 7 y 8.
- Diferencias entre la temporada 4 y las temporadas 6, 7 y 8.

4.0.5. Correlación de Spearman

La cantidad de dinero que gana un jugador de la NBA va cambiando a medida que pasan las temporadas. Es muy poco usual que los salarios permanezcan constantes con el paso de los años. Estos salarios pueden cambiar por varios motivos, siendo el más común el traspaso de un jugador a otro equipo. Sin embargo así los salarios cambien cada temporada, los valores salariales suelen estar muy relacionados entre temporadas. Por tal motivo es razonable hacer un test de correlación entre los datos repetidos de cualesquiera dos temporadas para verificar esto.

Como tenemos datos de 8 temporadas de la NBA, se pueden escoger 28 parejas de temporadas para calcular su correlación. En primer lugar se va a calcular la correlación entre las temporadas 2011-2012 y 2012-2013. La temporada 2011-2012 cuenta con el salario de 464 jugadores, y la temporada 2012-2013 cuenta con el salario de 478 jugadores. Se realiza una unión de estas dos temporadas para conocer cuantos jugadores se encuentran con salario en ambas.

```
XI_XIII<- sqldf("select XI_XIIc.player,
              XI_XIIc.team as teamXI_
                 XII,
              XII_XIIIc.team as teamXII
                 _XIII,
              XI_XIIc.salary_cc as
                 salaryXI_XII,
              XII_XIIIc.salary_cc as
                 salaryXII_XIII
            FROM XI_XIIc
            INNER JOIN XII_XIIIc
            ON XI_XIIc.player = XII_XIIIc
              .player")
> length(XI_XIII$player)
[1] 385
```

Al realizar la unión se puede observar que 385 jugadores tenían un contrato activo durante las temporadas 2011-2012 y 2012-2013. Además ahora tenemos los valores salariales de ambas temporadas en un mismo marco de datos. A continuación para los dos vectores que contienen los valores salariales se realizarán pruebas de normalidad para saber qué test de correlación debemos utilizar.

```
> ks.test(XI_XIII$salaryXI_XII, "pnorm",
+        mean(XI_XIII$salaryXI_XII),sd(
   XI_XIII$salaryXI_XII))

  One-sample Kolmogorov-Smirnov test

data:  XI_XIII$salaryXI_XII
D = 0.16896, p-value = 5.684e-10
alternative hypothesis: two-sided

> ks.test(XI_XIII$salaryXII_XIII, "pnorm"
   ,
+        mean(XI_XIII$salaryXII_XIII),sd
   (XI_XIII$salaryXII_XIII))

  One-sample Kolmogorov-Smirnov test

data:  XI_XIII$salaryXII_XIII
D = 0.17302, p-value = 1.952e-10
alternative hypothesis: two-sided
```

Como se rechazó la hipótesis de normalidad en los dos vectores, se recomienda realizar un test de correlación de Spearman para verificar que la correlación es mayor que 0. Además se calcula la correlación estimada entre los dos vectores de valores.

```
> cor.test(XI_XIII$salaryXI_XII,XI_XIII$
   salaryXII_XIII, conf.level = 0.95,
+          method="spearman", exact=FALSE
   , alternative="greater")

  Spearmans rank correlation rho

data:  XI_XIII$salaryXI_XII and XI_XIII$
   salaryXII_XIII
S = 1577563, p-value < 2.2e-16
alternative hypothesis: true rho is
   greater than 0
sample estimates:
     rho
0.8341334
```

Se observa que se rechaza la hipótesis $H_0 : \rho = 0$ para las temporadas 2011-2012 y 2012-2013 dado que el valor p calculado es menor que $0{,}05$. Ahora este proceso se repite otras 27 veces para poder realizar el test de correlación con las 28 parejas posibles de temporadas. A continuación se presentan en diferentes tablas los resultados.

Temporada	11/12	12/13	13/14	14/15	15/16	16/17	17/18	18/19
11/12	-	385	332	284	251	222	187	150
12/13	-	-	390	328	292	258	218	184
13/14	-	-	-	403	346	310	265	224
14/15	-	-	-	-	403	349	304	253
15/16	-	-	-	-	-	416	357	303
16/17	-	-	-	-	-	-	439	362
17/18	-	-	-	-	-	-	-	449

Cuadro 2: Jugadores con contrato activo en ambas temporadas

Se evidencia que a medida que hay más distancia de tiempo entre las temporadas, existen menos jugadores en común. Esto puede deberse a que muchos jugadores terminaron su contrato y no siguieron compitiendo en la NBA, o que los jugadores más veteranos se retiraron y dejaron de competir en la liga. A todas las columnas de salarios se les realiza el test de normalidad de Kolmogorov-Smirnov, y en todos los casos se rechazo la hipótesis de normalidad con un nivel de significancia del $99{,}9\,\%$.

Ahora se quiere probar la hipótesis $H_0 : \rho = 0$ vs $H_a : \rho > 0$. Como los datos no parecen seguir una distribución normal entonces para probar esta hipótesis se tendrá en cuenta el test de correlación de Spearman. A continuación se presenta el valor p de todos los tests realizados a cada pareja de temporadas.

Temporada	11/12	12/13	13/14	14/15	15/16	16/17	17/18	18/19
11/12	-	0	0	0	0	0	0.4	0.32
12/13	-	-	0	0	0	0	0.001	0.01
13/14	-	-	-	0	0	0	0	0
14/15	-	-	-	-	0	0	0	0
15/16	-	-	-	-	-	0	0	0
16/17	-	-	-	-	-	-	0	0
17/18	-	-	-	-	-	-	-	0

Cuadro 3: Valor p de test de correlación

Se observa que en casi todos los casos se rechaza la hipótesis nula de correlación igual a 0. Ahora con la misma función, calculamos la correlación estimada entre los datos de las dos temporadas. A continuación se presentan los resultados.

Temporada	11/12	12/13	13/14	14/15	15/16	16/17	17/18	18/19
11/12	-	0.83	0.67	0.52	0.27	0.23	0.13	0.04
12/13	-	-	0.84	0.68	0.41	0.31	0.20	0.17
13/14	-	-	-	0.85	0.61	0.49	0.31	0.3
14/15	-	-	-	-	0.8	0.56	0.46	0.34
15/16	-	-	-	-	-	0.78	0.61	0.46
16/17	-	-	-	-	-	-	0.83	0.68
17/18	-	-	-	-	-	-	-	0.88

Cuadro 4: $\hat{\rho}$ para cada pareja de temporadas

5. Conclusiones

- Los salarios de los jugadores de la NBA entre las temporadas 2011-2012 y 2018-2019, no cumplen el supuesto de normalidad, dado que tiene una asimetría positiva. En este contexto, esto se debe a que existen varios jugadores con salarios extremadamente altos por sus excelentes actuaciones deportivas. Estos son los jugadores que anotan más puntos, consiguen más rebotes, realizan más asistencias, y contribuyen en su equipo con más victorias.
- El cambio de patrocinio de Adidas a Nike generó un aumento en el salario de los jugadores de la NBA, de la temporada 2014-2015 a la temporada 2015-2016.
- Existe una relación entre las medianas de los salarios en las diferentes temporadas cuando las temporadas son dirigidas por un mismo patrocinador, sin embargo, cuando existen cambios de patrocinio se evidencia diferencias entre las medianas de los salarios de los jugadores.
- La correlación es muy alta entre temporadas consecutivas, y a medida que va pasando el tiempo estas correlaciones van disminuyendo. Esto puede deberse a que el aumento en los salarios de los jugadores entre temporadas consecutivas es algo usual, y solo a unos pocos jugadores (en especial a los veteranos) se les reduce el sueldo. A medida que va pasando el tiempo el sueldo va variando considerablemente y a unos jugadores que antes les aumentaba constantemente, ahora se les disminuye hasta su retiro. Sería interesante observar como se comporta el salario con la edad de los jugadores, ya que al parecer esta tiene una relación muy alta.

Referencias

[1] *NBA Salaries. (s. f.)*. hoopshype. Recuperado 17 de mayo de 2021, de https://hoopshype.com/salaries/

[2] Lago, Á. J. (2018, 21 enero). *Salarios NBA: los jugadores más y menos rentables de la liga.* Sweet Hoops. https://www.sweethoops.com/salarios-nba-jugadores-rentables/

[3] Melo O. (2020) "Análisis de Regresión [Notas de Clase y Código R]" *Universidad Nacional de Colombia sede Bogotá*, Bogota D.C.

[4] *Baloncesto. (s.f)*. "Baloncesto deporte como entretenimiento". Recuperado de: https://www.fbrm.es/reglamento/

[5] Rovell D. (2015) "NBA signs deal with Nike; logo to appear on uniforms" *ESPN*. Recuperado de: https://www.espn.com/nba/story/id/13053413/nba-signs-8-year-apparel-deal-nike.

[6] *¿Cómo funciona la NBA? Tope salarial, traspasos, contratos y draft. (s. f.)*. medium. Recuperado 16 de mayo de 2021, de https://https://medium.com/@kanelonmonger/c %C3 %B3mo-funciona-la-nba-tope-salarial-traspasos-contratos-y-draft-f025c01c8064

[7] *NBA Salary Cap FAQ. (s. f.)*. Wayback Machine. Recuperado 16 de mayo de 2021, de https://https://web.archive.org/web/2002040817 1815/http://members.cox.net/lmcoon/salarycap .htm

[8] *Historia de la NBA. (s. f.)*. Wikipedia. Recuperado 16 de mayo de 2021, de https://https://es.wikipedia.org/wiki/Historia _de_la_NBA