

UNIVERSIDAD NACIONAL DE ROSARIO
Facultad de Ciencias Económicas y Estadística
Licenciatura en Estadística

TRABAJO PRÁCTICO "ESTADÍSTICA EN LA NBA"

Materia: Análisis de Regresión
Año: 2023



Docentes:
Noelia Castellana
Luciana Magnano
Ayudante alumno:
Julián L'heureux

I. Introducción

La NBA, National Basketball Association, es sin dudas la mejor liga de baloncesto profesional del mundo que se disputa en Estados Unidos desde el año 1949. Está conformada por 30 equipos, de los cuales 29 están radicados en Estados Unidos y un sólo equipo es oriundo de Canadá (Toronto Raptors). Estos equipos se dividen en 2 conferencias, la Este y la Oeste, compuestas por 15 equipos en cada una. A su vez, cada conferencia está compuesta por 3 divisiones, que agrupan 5 equipos cada una.



En la NBA, las estadísticas desempeñan un papel fundamental tanto para los equipos como para los aficionados a la hora de evaluar el rendimiento de los jugadores, analizar estrategias y tomar decisiones ya que proporcionan una visión detallada del desempeño individual y colectivo de los jugadores en diversos aspectos del juego.

Estas estadísticas permiten medir y comparar aspectos clave del juego como ser: los puntos anotados, los rebotes capturados, las asistencias repartidas, los robos de pelota y los bloqueos realizados. Estas métricas ayudan a los entrenadores y gerentes a evaluar la eficiencia y el impacto de un jugador en los partidos.

Además, las estadísticas en la NBA también son ampliamente utilizadas por los aficionados y los medios de comunicación para seguir y analizar el juego. Los fanáticos pueden utilizarlas para evaluar a sus jugadores y equipos favoritos, realizar pronósticos, participar en competiciones como el "fantasy basketball" y enriquecer su experiencia general en la NBA.

Para conocer más sobre la NBA les compartimos los siguientes links:

<https://spain.id.nba.com/noticias/nba-que-es-como-funciona>

<https://entrenadordebasquet.com/nba/>

Dentro de este contexto, un grupo de periodistas de ESPN decide escribir un artículo sobre los factores que influyen en el número de puntos totales que un jugador anota en una temporada.

Se realizaron las siguientes preguntas:

- ¿A mayor altura del jugador, mayor cantidad de puntos anotados?
- ¿La posición de juego influye en los puntos anotados?
- ¿Cómo influye la edad en la cantidad de puntos anotados?
- ¿Influye en los puntos anotados la cantidad de minutos que el jugador estuvo en el campo de juego?
- ¿La performance histórica del jugador influye en el total de puntos anotados en esa temporada?

II. Data sets

Para poder investigar sobre este tema los periodistas exploraron información disponible que contaban para una temporada determinada. Cuentan con dos conjuntos de datos:

- **Nba_año.csv:** contiene métricas por jugador durante la temporada seleccionada (año).
- **Players.csv:** contiene características de todos los jugadores.

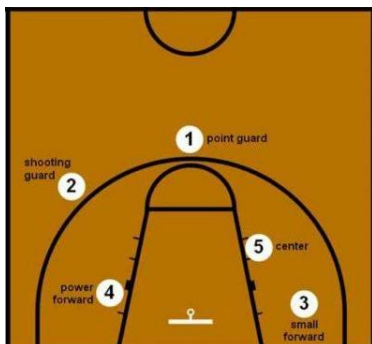
Estructura del archivo "Nba_año.csv:"

En la primera fila de este archivo se encuentran los nombres de las variables. A partir de la segunda fila, cada una de ellas corresponde a un jugador.

En las columnas se encuentran los datos de las características relevadas que se describen a continuación:

Columna	Descripción
Player	Nombre del jugador. Identificador único del jugador.
Year	Año que corresponde a la temporada.
Age	Edad del jugador en ese año.
Tm	Equipo para el cual jugó el jugador ese año. NO se utiliza en los análisis.
Pos	Posición de juego del jugador: <ul style="list-style-type: none">- PG: Point Guard.- SG-SF: Shooting Guard / Small Forward- PF-C: Power Forward / Center
G	Cantidad total de partidos jugados.
MP	Cantidad total de minutos jugados.
perf	Indicadora sobre la performance histórica del jugador <ul style="list-style-type: none">- 1 : muy buena performance- 0 : performance buena/regular
PTS	Cantidad total de puntos anotados.

En la siguiente figura se presentan las 5 posiciones en la cancha de básquet:



Estructura del archivo "Players.csv:"

En la primera fila de este archivo se encuentran los nombres de las variables. A partir de la segunda fila, cada una de ellas corresponde a un jugador.

En las columnas se encuentran los datos de las características relevadas que se describen a continuación:

Columna	Descripción
Player	Nombre del jugador. Identificador único del jugador.
Height	Altura del jugador.
Weight	Peso del jugador (promedio para todas las temporadas).

Estos data set fueron extraídos del repositorio [www.kaggle.com](https://www.kaggle.com/drgilermo/nba-players-stats) mediante la siguiente ruta: <https://www.kaggle.com/drgilermo/nba-players-stats>

Unificación de data sets:

Para poder contar con un único data set deberá poder incorporar la información de la altura y del peso de los jugadores que está en el data set "Players.csv" al data set "Nba_año.csv". La clave de vinculación de ambos archivos es el campo "Player".

El data set final contará con la misma cantidad de jugadores que contaba el data set Nba_año.csv con el agregado de las 2 columnas nuevas.

III. Preguntas

1) ¿Qué características presentan los jugadores estudiados?

- 1.1 Realice un análisis descriptivo univariado de las variables relevadas. ¿Hay valores atípicos? ¿Hay valores perdidos?
- 1.2 ¿Qué relación tienen las variables consideradas explicativas con los puntos totales anotados? ¿Entre ellas presentan alguna relación?

2) Modelo 1

En una primera instancia se desea estudiar la relación entre el total de minutos jugados y los puntos totales anotados. Para ello se decide ajustar un modelo de regresión lineal simple (Modelo 1).

- 2.1 Escribir la ecuación y los supuestos para el Modelo 1.
- 2.2 ¿Existe regresión? Plantear la hipótesis correspondiente, escribir la estadística de prueba y su distribución y concluir en términos del problema.
- 2.3 Escribir la ecuación estimada e interpretar los parámetros estimados.
- 2.4 Comentar acerca del valor del coeficiente de determinación
- 2.5 Evaluar si se cumplen los supuestos del modelo. ¿Hay observaciones atípicas? ¿Corresponde a algún jugador particular?

3) Modelo 2

En una segunda instancia se desea estudiar la relación entre todas las variables relevadas y el total de puntos anotados. Para ello se decide ajustar un modelo de regresión lineal múltiple (Modelo 2).

- 3.1 Escriba la ecuación y los supuestos para el Modelo 2. Incorporar la variable dicotómica performance. No incorpore interacciones.
- 3.2 ¿Existe regresión? Plantee las hipótesis correspondientes, arme el cuadro ANOVA e interprete en términos del problema.
- 3.3 Estime los coeficientes del modelo y realice los test parciales. ¿Qué sugieren estos test?
- 3.4 ¿Existe multicolinealidad?

4) Modelo 3

Según el análisis de multicolinealidad y el resultado de los test parciales plantee un nuevo modelo llamado Modelo 3. Incorporar la variable dicotómica performance. Evaluar incorporar interacciones

4.1 Estime el nuevo Modelo 3 y evalúe si se cumplen los supuestos del modelo mediante un análisis de residuos. ¿Hay alguna observación atípica? Realice los gráficos de regresión parcial. Comente.

4.2 Si no se cumplen los supuestos plantear y realizar transformaciones para intentar solucionar el problema.

5) Modelo final

Luego de lo concluido en los puntos anteriores arribar a un Modelo final.

5.1 Escriba la ecuación estimada del Modelo final en forma general y según la variable performance. Interpretar los coeficientes en términos del problema.

5.2 Elija un coeficiente e interprete su intervalo de confianza del 95%.

5.3 Responda a las preguntas que se realizaron los periodistas según el modelo final estimado.

5.4 El entrenador del equipo Miami Heat quiere utilizar el modelo para estimar la cantidad de puntos que va a anotar al final de la temporada 2023 un jugador nuevo llamado Nicolas Sánchez. Este jugador mide 1.95mts, pesa 96 kg., su posición de juego es PG (Point Guard), presentó una performance regular en las temporadas previas y se cree que va a jugar 29 partidos con un total de 510 minutos. Presentar la estimación puntual y por intervalo del 95%.

6) Informe periodístico

Los periodistas de ESPN nos solicitaron escribir un breve texto con los resultados que obtuvimos con nuestro modelo para que sea leído por toda la audiencia. Además, sugirieron que incluyamos en el informe si hay jugadores para los cuales se observó un comportamiento diferente del resto.