

Análisis de regresión sobre la relación de las estadísticas de MPG y PPG en partidos de postemporada de la NBA

Rudy Miranda Bastias

Github (scripts): <https://github.com/SolaireLordOfSunlight/Linear-Regression>

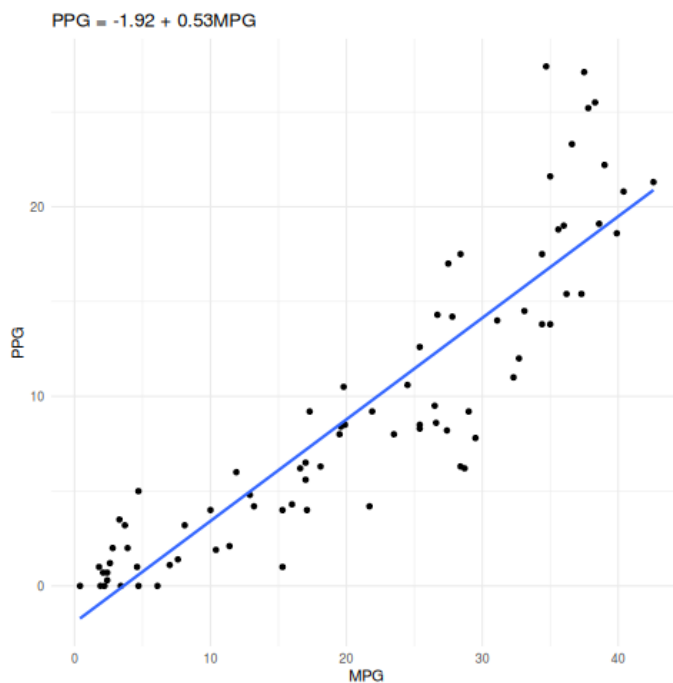
Abril, 2023

Introducción

Se busca confirmar si la relación entre la cantidad de puntos anotados por un jugador de la NBA, frente a la cantidad de minutos jugados por partido tienen una relación del tipo lineal.

Los datos corresponden a los playoff (post-temporada) de la temporada 2021-22 de la NBA obtenidos de su sitio web oficial.

Un punto a enfatizar es que las 79 unidades de observación son jugadores de la misma posición, donde se toma en cuenta su promedio de puntos por partidos (PPG) y su promedio de minutos por partido (MPG).



Modelo Poblacional

$$PPG = \beta_0 + \beta_1 MPG + \varepsilon \quad (1)$$

Estimación de Parametros

Comandos R

```
table_values <- read.csv("./nba_stats.csv")  
(summary(lm(formula = PPG ~ MPG, data = table_values)))
```

Resultados

		$\hat{\beta}_0$	$\hat{\beta}_1$	$R^2 = 0,8037$
PPG	=	-1.92	0.53	
s.e		0.73	0.03	
t		-2.65	17.9	
valor-p		0.01	≈ 0	

Aquí el $\hat{\beta}_1$ indica que los jugadores en promedio anotan 0.53 por cada minuto jugado. Por otro lado $\hat{\beta}_0 = -1,92$ es un número que no tiene sentido en la realidad, sería buena idea quitar el intercepto.

Interpretando R^2 , el 80 % de la variación de la estadística *PPG* se explica por la de *MPG*.

Validación del Modelo

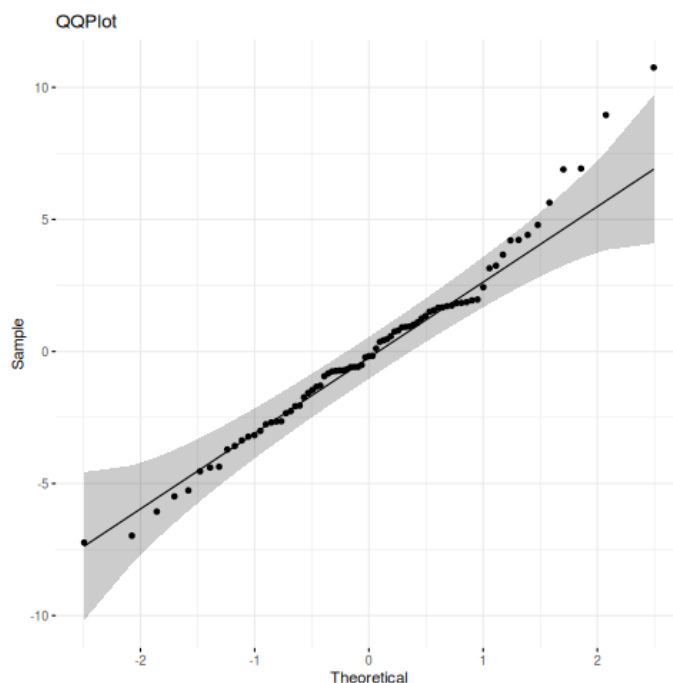
Proponemos las hipótesis

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

rechazando H_0 dado el valor- $p \approx 0$ del resumen anterior.

Análisis de Residuos



Una hipótesis que deben cumplir nuestro modelo es que $\varepsilon_i = (Y_i - \bar{Y}) \sim N(0, \sigma^2)$

Aplicando distintos test de normalidad, tanto paramétricos como no paramétricos

Test	Valor P
Jarque-Bera	0.04469
Kolmogorov-Smirnov	3.847e-05
Shapiro-Wilk	0.1174
Anderson-Darling	0.1493

Con un nivel de significancia de 0.05, podríamos considerar la distribución como una normal.

Ahora, al hacer un test de hipótesis con el comando *t.test* sobre la media con la hipótesis nula $H_0 : \mu = 0$ obtenemos un valor $p = 1$, con lo que aceptamos H_0 .

Con ello se cumplen los dos supuestos de la distribución de los residuos.

V-A. *Independencia de los Residuos*

Aplicando el test de Durbin-Watson, con las siguientes hipótesis

$$H_0 : \text{los errores son independientes}$$
$$H_1 : \text{los errores no son independientes}$$

Aceptamos la hipótesis nula, puesto que el valor- $p = 0,21 > 0,05$ el nivel de significancia con el cual trabajamos en este informe.

Referencia: https://fhernanb.github.io/libro_regresion/indep.html

Conclusión

Es clara la relación entre los minutos jugados y los puntos anotados, pero sería una buena decisión incluir mas variables a este modelo en vez de solo dejarlo en dos, y como se dijo anteriormente podríamos quitar el intercepto de nuestra función de regresión poblacional ya que al no jugar no se puede anotar una cantidad negativa de puntos.