## **Proyecto de Estadística Curso 2023**

2022 MLB Player Stats – Variante 3

### **Conjunto de datos**

2022 MLB Player Stats – Batting

### **Ejercicio 1**

Analice con qué variables existe relación lineal o monótona con la variable ‘HR’ (Home runs hit/allowed), para los jugadores de la Liga Nacional que hayan hecho más de 25 Doble Hit (‘2B’).

a) Si tomamos ‘HR’ como variable dependiente seleccione 2 variables independientes utilizando el método ‘hacia adelante’ (forward) para hacer una regresión lineal múltiple. Apóyese de un software matemático para hacer la regresión.

b) De la regresión anterior compruebe los supuestos. El modelo es válido?

c) Ahora seleccione como variables independientes ‘RBI’ (Runs batted in) y ‘SO’ (Strikeouts) y haga una nueva regresión múltiple. Compare este modelo con el del inciso ‘a’ y diga cuál de los dos es mejor matemáticamente.

### **Ejercicio 2**

Haga un análisis exploratorio del comportamiento por equipo, de la variable ‘H’ (Hits/hits allowed) de sus jugadores, solamente de los equipos ‘WSN’, ‘CHC, ‘ATL’ y ‘COL’ de la Liga Nacional (‘NL’), que hayan robado al menos una base (‘SB’ Stolen bases). Apóyese de histogramas y cajas de bigote.

a) Las medias de Hits por equipos, de los jugadores que al menos robaron una base, solo seleccionando los equipos ‘WSN’, ‘CHC, ‘ATL’ y ‘COL’ de la Liga Nacional (‘NL’), se pueden considerar estadísticamente iguales? (Recuerda comprobar los supuestos)

b) Haga una reducción de dimensión sobre las medias de las características agrupadas por todos los equipo de la Liga Nacional, usando árbol jerárquico. Agrupe en 4 clústeres

### **NOTA:**

Para cada test que utilice considere un nivel de confianza del 95%