## **Proyecto de Estadística Curso 2023**

2022 MLB Player Stats – Variante 1

### **Conjunto de datos**

2022 MLB Player Stats – Pitching

### **Ejercicio 1**

Analice con qué variables existe relación lineal o monótona con la variable ‘BB’ (Bases on balls/walks).

a) Si tomamos ‘BB’ 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. Cuál no se cumple? Cómo se interpreta esto geométricamente?

c) Si ahora nos restringimos a los datos del equipo ‘TOT’ de la liga ‘MLB’, utilizando igualmente ‘BB’ como variable dependiente y las 2 variables independientes escogidas en el ejercicio anterior, seleccione una nueva variable que nos garantice el cumplimiento de los supuestos y un buen modelo de regresión múltiple.

### **Ejercicio 2**

Haga un análisis exploratorio del comportamiento por equipo, de la variable ‘SO/W’ (Strikeouts / Wins) de sus jugadores, solamente de la Liga Americana (‘AL’). Apóyese de histogramas y cajas de bigote.

a) Las medias de la proporción ‘SO/W’ de los equipos ‘OAK’, ‘SEA’, ‘TEX’, y ‘NYY’ se pueden considerar estadísticamente iguales? (Recuerda comprobar los supuestos)

b) Haga una reducción de dimensión sobre las características de los jugadores que participaron en la Liga Americana, usando K-Means. Agrupe en 4 clústeres.

### **NOTA:**

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