

Introducción a la Ciencia de Datos y Machine Learning

Tarea : Simulación Bootstrap

Profesor: Jimmy Hernández Morales
Ayu. Antonio Soriano Flores

En los problemas solicitados, se requiere entregar código y salida del código en un PDF.

1. La base *Fubol.csv* contiene la información de $n = 8698$ partidos del fútbol español donde la columna *GL* se refiere a lo número de goles que anotó el equipo local mientras que la columna *GV* corresponde a los goles anotados por el equipo visitante.

Con la tabla anterior, construya las variables $X :=$ goles anotados en el partido y la variable $Z :=$ diferencia absoluta de goles entre los dos equipos.

Con esta información realice lo siguiente:

- Asumiendo que $X \sim \text{Pois}(\lambda)$ y utilizando la técnica del Bootstrap paramétrico, encuentre un intervalo al 85 % de confianza para $\mathbb{P}(X > 5)$, es decir encuentre un intervalo de confianza al 85 % para la probabilidad de que un partido en el fútbol español supere los 5 goles. (Realice al menos 50,000 simulaciones bootstrap)
- Suponga ahora que $X \sim F_X(x)$ con $F_X(x)$ una distribución desconocida, conteste la pregunta anterior utilizando la técnica de bootstrap no paramétrico. (Realice al menos 50,000 simulaciones bootstrap).
- Suponga que $Z \sim F_Z(z)$ con $F_Z(z)$ una distribución desconocida. Encuentre un intervalo al 99 % confianza para la probabilidad de que un partido del fútbol español tenga un empate. (Realice al menos 50,000 simulaciones bootstrap).
- Asumiendo que $Z \sim \text{Pois}(\lambda)$, encuentre un intervalo al 80 % de confianza para la probabilidad de que el equipo F.C. Barcelona gane por mas de 2 goles como visitante.(Realice al menos 50,000 simulaciones bootstrap).
- Repita los problemas anteriores pero paralelizando su código (Realice al menos 200,000 simulaciones bootstrap).