



---

# Análisis estadístico sobre una base de datos de béisbol.

---

**Miguel Cordero**  
Universidad Simón Bolívar  
Caracas, Venezuela  
[15-10326@usb.ve](mailto:15-10326@usb.ve)

**Eduardo Gavazut**  
Universidad Simón Bolívar  
Caracas, Venezuela  
[13-10524@usb.ve](mailto:13-10524@usb.ve)

**Luis Riera**  
Universidad Simón Bolívar  
Caracas, Venezuela  
[16-10976@usb.ve](mailto:16-10976@usb.ve)

8 de abril de 2022

**RESUMEN:** Este es un ejemplo de plantilla generado en RStudio para mostrar como se vería el proyecto final. Se realizarán una serie de análisis cuantitativos y cualitativos sobre una base de datos de béisbol. Se exponen los resultados y algunas conclusiones que se pueden extraer de los mismos. Cualquier error puede ser notificado para su corrección final. ESTO NO ES UN RESUMEN, se los estoy diciendo literal XD.

**Palabras clave:** Proyecto, Estadística, Rstudio, Beisbol

---

---

## 1. ANOVA PARA LA TASA DE BATEO

---

Para realizar el analiza de varianza sobre la variable  $X_1$  O la tasa de bateo. Primero dividimos los datos en 3 grupos:

- Grupo 1: los bateadores con una tasa de bateo igual a  $(X_1 < 0,200)$ .
- Grupo 2: los bateadores con una tasa de bateo igual a  $(0,200 \leq X_1 < 0,300)$ .
- Grupo 3: los bateadores con una tasa de bateo igual a  $(0,300 \leq X_1)$

Con esta agrupación se decidió por considerar un análisis de varianza con bloques aleatorizados donde los bloques serán los grupos y los tratamientos o métodos las distintas variables de la base de datos.

Con la tabla 1 podemos apreciar las medias de los valores agrupados. Con esto valores, se puede aplicar el comando `anova` de R para obtener la tabla ANDEVA detallada en la tabla 2, donde se obtiene que el p-valor para lo grupos es de 0,6198 que es alto, indicando que la hipótesis nula para los grupos no se puede rechazar por lo que las medias por grupos son iguales. Sin embargo, para las medias por variable o método se obtuvo un p-valor de 0,0004 que es bastante bajo, incluso significativo indicando que las medias son distintas tal como se esperaba por los datos analizados.

Con esto podemos afirmar que los promedios de las tasas de las otras variables son iguales por cada grupo.



**Tabla 1:** Tabla a dos factores para las medias de las variables

	X1	X2	X3	X4	X5	X6
Grupo1	0.1935000	0.0820000	0.0365000	0.0070000	0.00650	0.1965000
Grupo2	0.2580400	0.1327600	0.0414400	0.0087200	0.02228	0.1204800
Grupo3	0.3212778	0.1837222	0.0542778	0.0153333	0.02900	0.0716667

**Tabla 2:** Tabla a dos factores para las medias de las variables

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
grupos	2	0.0019826	0.0009913	0.5020692	0.6197546
variables	5	0.1334024	0.0266805	13.5132135	0.0003524
Residuals	10	0.0197440	0.0019744	NA	NA