

Se presenta una base de datos que recopila información de diferentes aplicaciones de la Play Store de Google. Dicha información presentada en variables que muestran atributos de las aplicaciones como el número de veces que esta ha sido instalada, el total de calificaciones que ha recibido y su distribución en las categorías que van desde una hasta cinco estrellas, etcétera.

Cuadro 1: Vista previa de algunas variables

Titulo	Total de calificaciones	Total de calificaciones cinco estrellas
Call of Duty®: Mobile - Season 4: Spurred & Burned	13572148	10501443
Mystic Messenger	419193	356875
Blockudoku® - Block Puzzle Game	414430	272911
City Driving 3D	272721	138933
Slugterra: Slug it Out 2	481615	412995

Considere la cantidad total de calificaciones como la covariable y a la cantidad de calificaciones en la categoría de cinco estrellas como la variable respuesta.

Su tarea como analista es realizar las siguientes tareas usando el software estadístico *R*.

1. Realice la lectura de la base de datos, posteriormente seleccione solo la covariable y la variable respuesta, filtre para solo quedarse con aquellas observaciones que tengan menos de 100000 calificaciones totales y pertenezcan a la categoría de juegos de acción. Finalmente guarde dichas observaciones en una nueva base de datos.
2. Elabore un gráfico de dispersión de los datos, luego de esto analícelo.
3. Escriba la ecuación del modelo de regresión, junto con sus supuestos. Ajuste un modelo de regresión lineal simple y añada la recta de regresión a la gráfica generada anteriormente.
4. Realice una interpretación del parámetro β_0 , ¿qué unidades tiene? Determine si este es significativo usando $\alpha = 0.05$ y si tiene sentido en el contexto de los datos.
5. Repita el proceso anteriormente enunciado con el parámetro β_1 .
6. Calcule un intervalo de confianza del 95 % para ambos parámetros del modelo. Antes de calcularlo responda, ¿dichos intervalos deberían contener al cero?
7. Use el modelo para predecir el número de calificaciones dentro de la categoría de 5 estrellas para las observaciones que no fueron usadas para ajustar el modelo. Realice un gráfico de los valores reales vs los valores predichos ¿qué observa en dicho gráfico?