
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

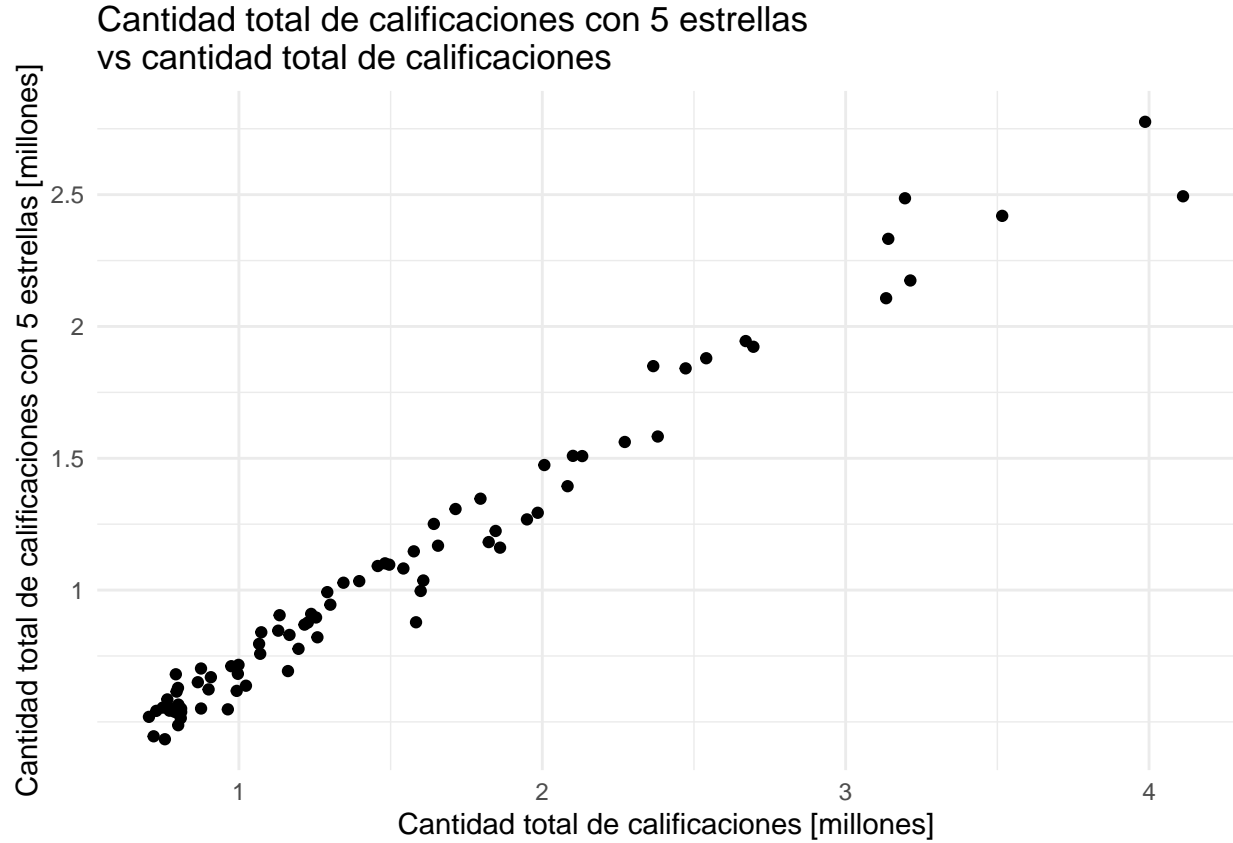
Considere la cantidad total de calificaciones como la covariable y a la cantidad de calificaciones en la categoría de cinco estrellas como la variable respuesta.

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

1. Realice la lectura de la base de datos, posteriormente filtre para solo quedarse con aquellas observaciones que tengan menos de 4121627 calificaciones totales y pertenezcan a la categoría de juegos de acción, seleccione solo la covariable y la variable respuesta. 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?

Solución

Ejercicio 2



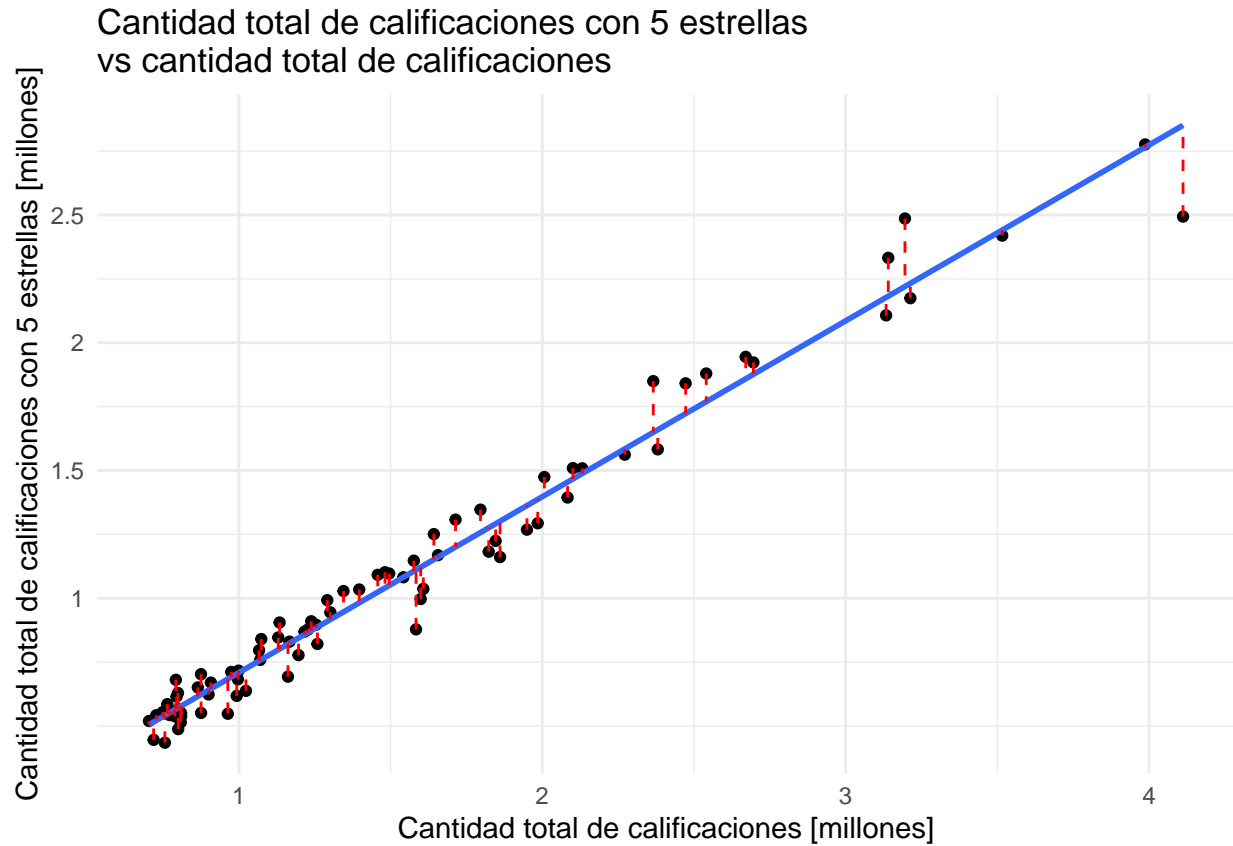
Se logra percibir una relación lineal entre ambas variables, donde estas están correlacionadas de manera positiva, es decir, a mayor número de calificaciones se tiene un crecimiento en las calificaciones con 5 estrellas

Ejercicio 3

A continuación se realiza la especificación del modelo de regresión.

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \epsilon_i \stackrel{iid}{\sim} N(0, \sigma), \quad 1 \leq i \leq 80$$

Adicionalmente se presenta el gráfico de dispersión con la respectiva recta de regresión.



Ejercicios 4 y 5

Cuadro 2: Resumen de los coeficientes

	Estimación	Error estándar	t_{0Obs}	P-valor
β_0	20303.78	22173.2195	0.9157	0.3627
β_1	0.6885	0.0128	53.7437	0

- β_0 no tiene interpretación y no es significativo tiene unidades de calificaciones de 5 estrellas.
- β_1 es la tasa de cambio de la cantidad de calificaciones de 5 estrellas por la cantidad de calificaciones totales, además es importante traer a colación que el parámetro es inferior a 1, característica que se espera (¿por qué?). Finalmente, el parámetro β_1 es significativamente distinto de cero (¿esto qué implica?).

Ejercicio 6

Previo a reportar los límites inferior y superior de los intervalos de confianza del 95 %, se hace mención de que se espera que el cero pertenezca al intervalo del intercepto mientras que de manera contraria, se espera que no pertenezca al intervalo de la pendiente.

Cuadro 3: Intervalos de confianza

Límite inferior	Límite superior
-23839.7091	64447.2691
0.663	0.714