

Los juegos más jugados de Steam

Jorge Rodríguez Fraile y Carlos Rubio Olivares

INTRODUCCIÓN Y OBJETIVOS

En este proyecto observaremos el comportamiento del jugador medio en la plataforma de videojuegos de Steam, la más utilizada del mundo.

Nuestro objetivo es analizar si lo más jugado es lo más valorado, si el precio influye en el número de jugadores, si existe homogeneidad en cuanto a los géneros, si los juegos actuales son los más jugados, o, en cambio, los más asentados y antiguos siguen liderando en cuanto a número de jugadores.

Hay que tener en cuenta que el mundo de los videojuegos es, en cierta manera, muy volátil, y está sujeta a un permanente cambio, por lo que estos datos se podrían quedar obsoletos en un corto período de tiempo.

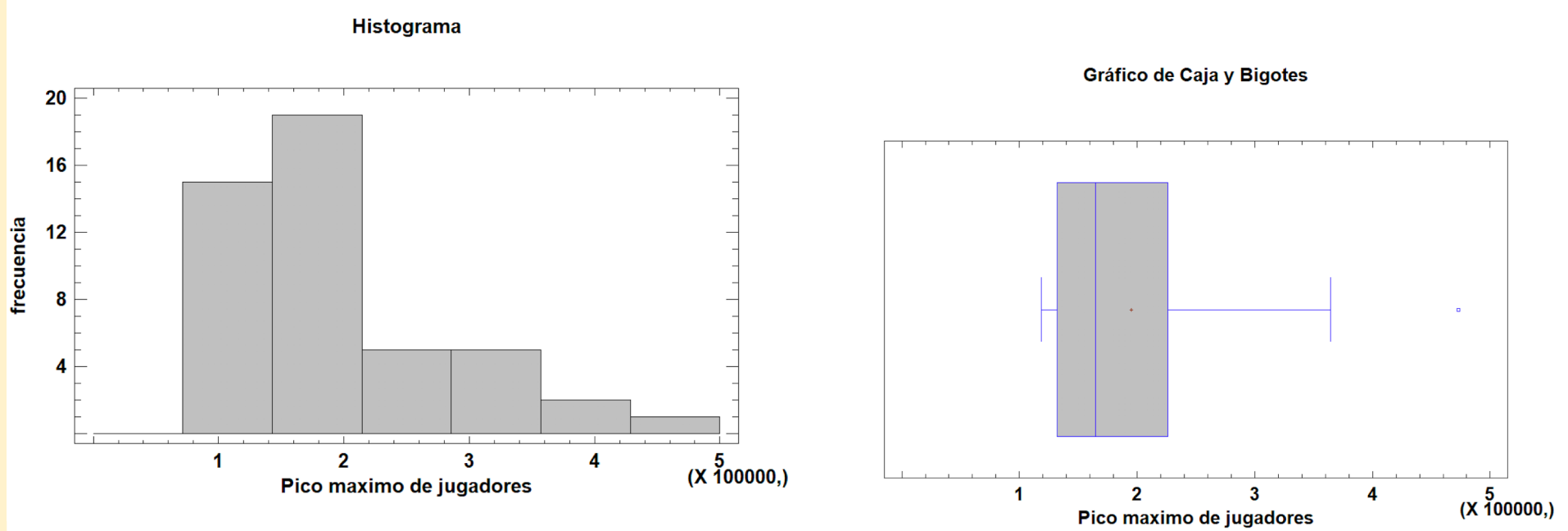
ESTUDIO DESCRIPTIVO

Tras la eliminación de los datos extremadamente atípicos obtenemos:

Media = 194.908 jugadores
Mediana = 164.688 jugadores
Intervalo modal es el de 178.571 de marca de clase.
Varianza = 6,774E9 jugadores²
Desviación típica = 82.308,7 jugadores
Coeficiente de variación de 42'23%

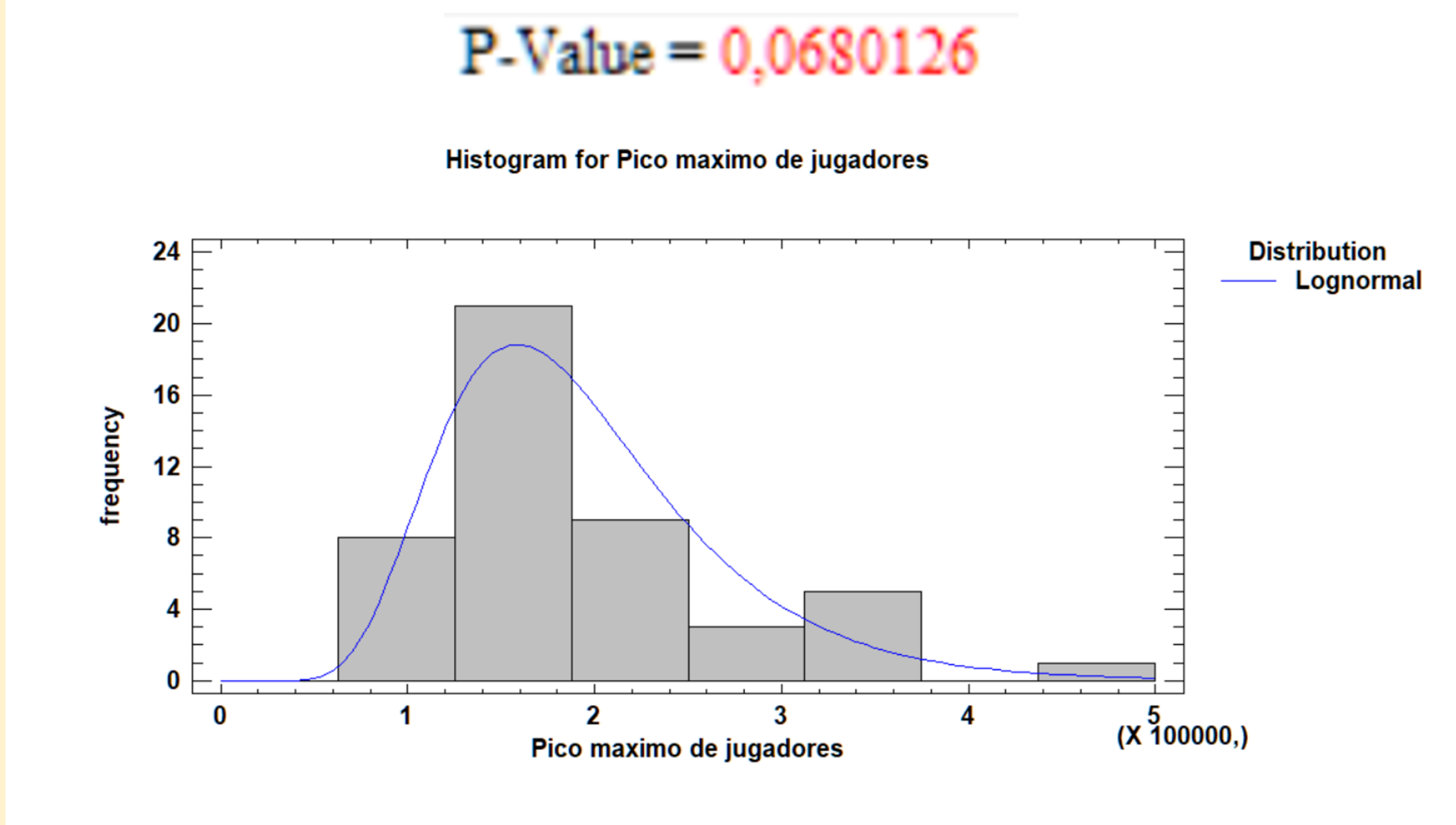
Asimetría positiva en lo que respecta a la forma, coeficiente de asimetría de 1'47.

Esto nos muestra que tenemos una concentración de datos en la parte izquierda de la muestra, lo que implica que hay muy pocos juegos con valores mayores a 300.000 jugadores, algo con bastante sentido.



AJUSTE DE DISTRIBUCIONES

La distribución que más se ajusta es la Lognormal para 8 clases: LogN(194032, 73677,8). Como hemos explicado en el apartado de descriptiva, hemos utilizado una lognormal para centrar los datos que estaban acumulados en la parte izquierda de nuestra muestra.



INTERVALOS DE CONFIANZA Y CONTRASTE DE HIPÓTESIS

El intervalo de confianza al 95% para la media se encuentra entre 170.741 y 219.074 jugadores, pero no podemos aplicar este estudio a nuestro contraste de hipótesis, ya que vamos a usar un contraste unilateral.

95,0% confidence interval for mean: 194908, +/- 24166,8 [170741.; 219074.]
95,0% confidence interval for standard deviation: [68396,5; 103379.]

La hipótesis que analizaremos será ver la media del pico máximo de jugadores de un juego de Steam será mayor de 200.000

Como el p-valor de este contraste de hipótesis es mayor a nuestro intervalo de confianza, (0,05) por lo que no podemos rechazar la hipótesis base y afirmar que la media es mayor a 200.000

Hypothesis Tests

Sample mean = 194032,
Sample standard deviation = 73677,8
Sample size = 50

95,0% lower confidence bound for mean: 194032, - 17469,0 [176563,]

Null Hypothesis: mean = 200000,
Alternative: greater than
Computed t statistic = -0,572766
P-Value = 0,715289
Do not reject the null hypothesis for alpha = 0,05.

REGRESIÓN MÚLTIPLE

Hemos realizado una regresión múltiple con la variable dependiente pico máximo de jugadores y como variables dependientes precio, tiempo de juego medio, valoraciones negativas, año de salida y valoraciones positivas. Al hacer este modelo, hemos observado que val pos y val neg tiene un p-valor mayor a 0,05, por lo que hemos intentado transformar nuestros datos con el log, pero no ha resultado efecto. Al final hemos optado por eliminar val pos y neg, (tras probarlas por separado), nuestra hipótesis es que muchas de estas valoraciones provienen de publicidad o gente que es extremadamente fan del propio juego, y crean numerosas valoraciones positivas con el mismo usuario. En cuanto al género, no poseemos suficientes datos como para hacer una relación con ellos. Por último, año de salida no aporta mucho, ya que hay juegos en el top de incluso 5 años de antiaüedad.

	Estimate	Standard Error	T	P-Value
Parameter				
CONSTANT	180151,	14833,4	12,1449	0,0000
Tiempo juego medio	2595,01	635,275	4,08486	0,0002
Precio	-1542,32	595,456	-2,59014	0,0130

Analysis of Variance					
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	8,77378E10	2	4,38689E10	8,62	0,0007
Residual	2,239E11	44	5,08863E9		
Total (Corr.)	3,11638E11	46			

Pico máximo de jugadores = 180151, + 2595,01*Tiempo juego medio - 1542,32*Precio

Pico máximo con un tiempo medio de 50 horas y precio de 60 euros: 217.363
Estimamos una rebaja de 10 euros y el mismo tiempo: sube 15.423 jugadores

Nuestra regresión tiene una R cuadrado de 28%, que es bastante poco, esto implica que nuestro modelo es bastante malo, nuestra explicación es que Steam es una plataforma cambiante constantemente, con más 130.000 de juegos en la actualidad, lo que implica que muchos de nuestros valores sean éxitos efímeros que en un corto plazo vuelvan a bajar de los tops.

CONCLUSIONES

Teniendo en cuenta todos los resultados previos, podemos afirmar que lo que de verdad influye en que un juego tenga muchos seguidores se debe principalmente al precio que tiene y el tiempo de entretenimiento que nos ofrece, por otro lado, la opinión externa es muy poco valorada. Debemos tener en cuenta que hemos quitado juegos muy famosos, debido a que la base de seguidores que tienen es enorme, pero aún así hemos podido observar cómo se comporta el usuario medio de esta plataforma y qué valores son los que más se tienen en cuenta. Por último, cabe recalcar que los éxitos temporales se han tenido en cuenta, como también grandes bajadas de precios en algunos juegos, por lo que muchos de ellos aparecen en el top 50 por estas causas, aún así, esto refuerza los resultados obtenidos, ya que estas características son tan importantes como para hacer que un juego suba tanto en los rankings.