

Estudio de la relación entre el precio de Tarjetas Gráficas y su rendimiento

Alvaro Pérez Vergara y Joan Carlos Naftanaila

INTRODUCCIÓN Y OBJETIVOS

En este estudio trataremos de analizar las diversas especificaciones y prestaciones de tarjetas gráficas sobre el precio de lanzamiento (USD), que será nuestra variable principal y trataremos de comparar esto con el estado de los precios en la actualidad.

Variables auxiliares:

- Rendimiento teórico (Tflops FP32)
- Rendimiento en tests (FPS)
- Ancho de banda de la gráfica (GB/s)
- Popularidad en steam (%)

Fuentes:

<https://videocardz.com/>

<https://gpu.userbenchmark.com>

<https://store.steampowered.com/hwsurvey/video-card/>

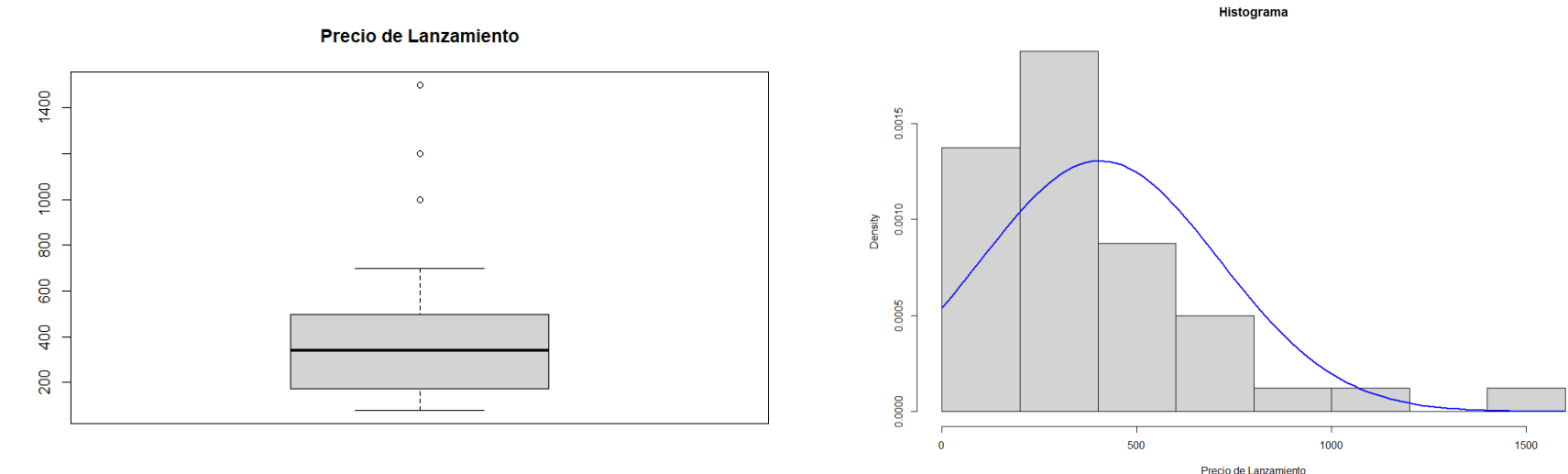
<https://www.tomshardware.com/news/gpu-pricing-index> (página de noviembre)

ESTUDIO DESCRIPTIVO

Presentamos las medidas de tendencia central y dispersión de nuestra variable principal de estudio en la siguiente tabla. “Precio de salida del producto”.

En el siguiente histograma del precio hemos decidido que el número óptimo de clases es 7, ya que es el número aproximado hacia arriba más próximo a \sqrt{n} (6,32). En él se observa una gran concentración de precios en torno a los intervalos de 0 a 400 dólares, que son los precios de la mayoría de gráficas orientadas al público general. A su vez, se pueden observar un par de valores atípicos por encima de 1200 dólares, que corresponden a las gráficas de más alta gama. Esta condición se observa sobre todo en el diagrama de caja. Gráficamente podemos observar que tenemos un par de datos atípicos-atípicos extremos

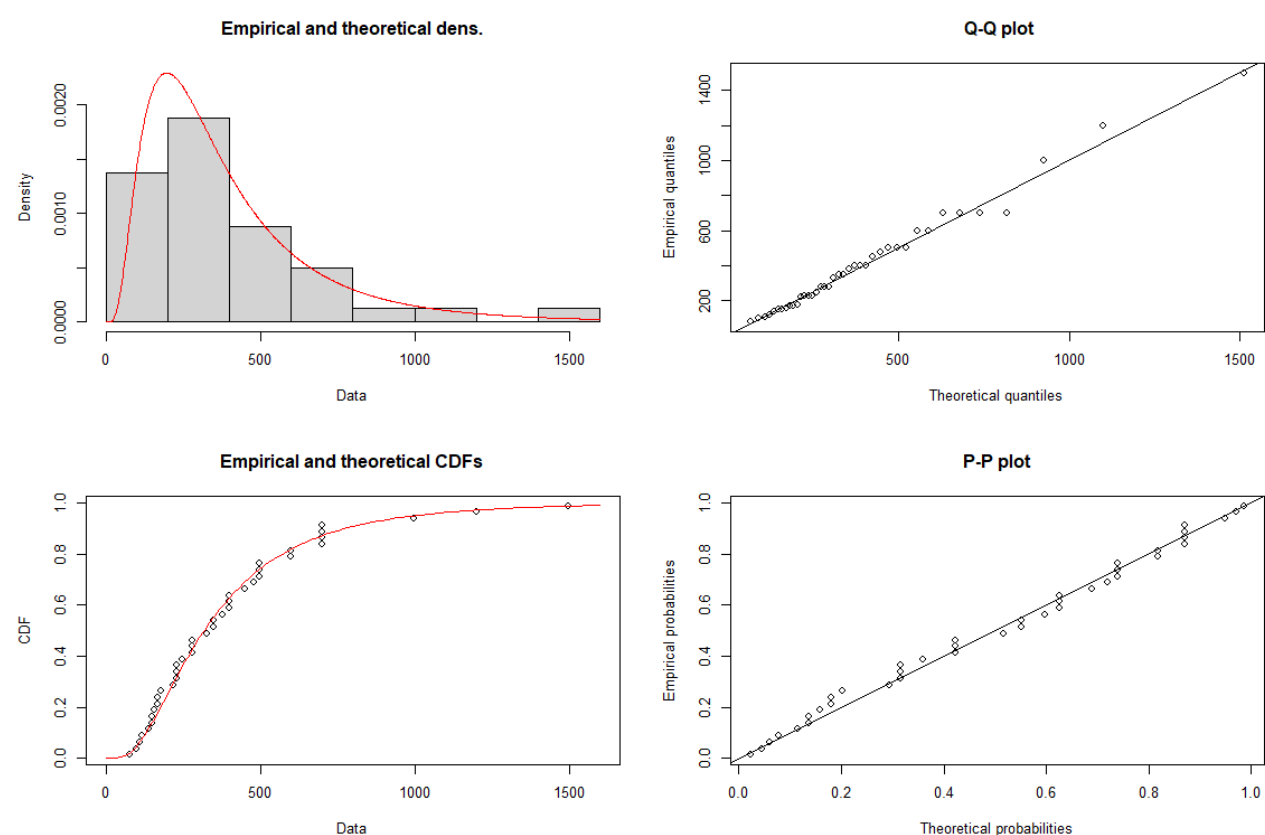
	Precio de lanzamiento (USD)
Nº de objetos	40
Mínimo	79
Máximo	1499
Rango	1420
Media	405.9230769
Mediana	329
Desv estándar	309.559567
Varianza	95827.12551
Coefficiente de variación	76.26064754
Cuartil 1	174
Cuartil 3	499
Rango intercuartílico	325
Coefficiente de asimetría	1.755624807
Curtosis	3.61597135



Como podemos observar estas variables no están normalizadas ya que presentan una clara asimetría. Aplicando una transformación logarítmica conseguimos corregirlas, mayoritariamente.

AJUSTE DE DISTRIBUCIONES

Al tratar de ajustar por el método de chi cuadrado nuestra variable precio con una distribución normal nos sale un p valor de 1.298578e-10, que es mucho menor a 0,05 por lo que podemos descartar este tipo de distribución. Tras aplicar una transformación logarítmica el resultado obtenido se ajusta con mayor precisión a la realidad tal y como muestra los siguientes gráficos:



Además, con este tipo de distribución nos sale un p valor de 0.9490561, que es mayor a 0,05, y por tanto, continuaremos con esta a lo largo del estudio.

INTERVALOS DE CONFIANZA Y CONTRASTE DE HIPÓTESIS

Para empezar, calculamos el intervalo de confianza de la media. Este intervalo tendrá una confianza del 95% ya que consideramos el valor $\alpha=0,05$.

Intervalo de confianza para la media: [311.0555; 500.4445].

Por otro lado, no podemos calcular el intervalo de confianza de la varianza. Esto se debe a que para ello se necesita que la muestra siga una distribución normal, y la nuestra se ajusta a una lognormal.

Para el contraste de hipótesis de nuestra población, tratamos de comparar los precios actuales de las gráficas con su precio de salida. La hipótesis planteada sería que las gráficas normalmente bajan de precio a cuanto más tiempo haya pasado de su salida, sin embargo actualmente son más caras que cuando salieron, ya que la minería de criptomonedas y la falta de materiales habrían encarecido los precios, es por esto que vamos a estudiar cómo se ha comportado la media del precio de las gráficas en este mercado. Siendo μ la media de los precios de lanzamiento y $\mu_1=937,4333$ la media de los precios actuales, y con un $\alpha=0,05$, tenemos que:

$H0: \mu \geq \mu_1$
 $H1: \mu < \mu_1$

```
One Sample t-test

data: Libros$`Precio de lanzamiento (USD)`
t = -11.005, df = 39, p-value = 7.941e-14
alternative hypothesis: true mean is less than 937.4333
95 percent confidence interval:
 -Inf 487.1537
sample estimates:
mean of x
405.75
```

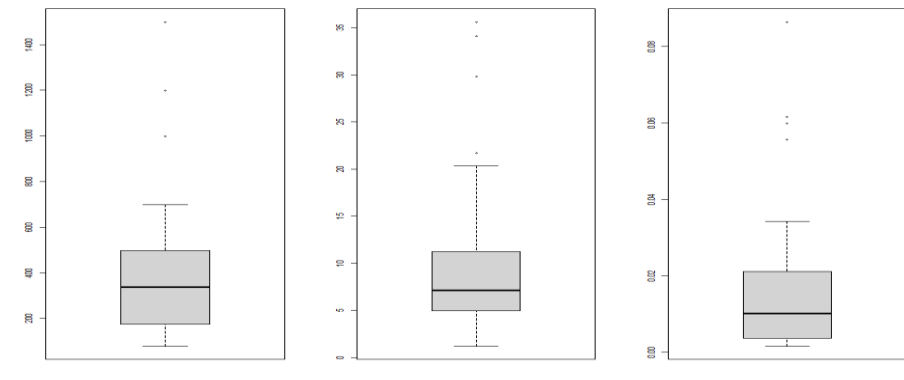
Aquí calculamos el p-valor:

```
> pnorm(-11.005)
[1] 1.80758e-28
```

Dado que el p-valor es menor a 0,05, rechazamos la hipótesis nula y podemos afirmar que el precio de las gráficas actualmente es mayor que su precio de salida.

REGRESIÓN MÚLTIPLE

Antes de hacer la regresión hemos analizado nuestros datos para detectar algunas tarjetas que pueden tener datos atípicos como la 3090 y la 3080 ya que son gráficas de gama entusiasta no orientadas al público general, vemos la discrepancia de los datos en la siguiente gráfica:



Comenzamos con el modelo inicial de regresión múltiple que combina todas las variables secundarias:

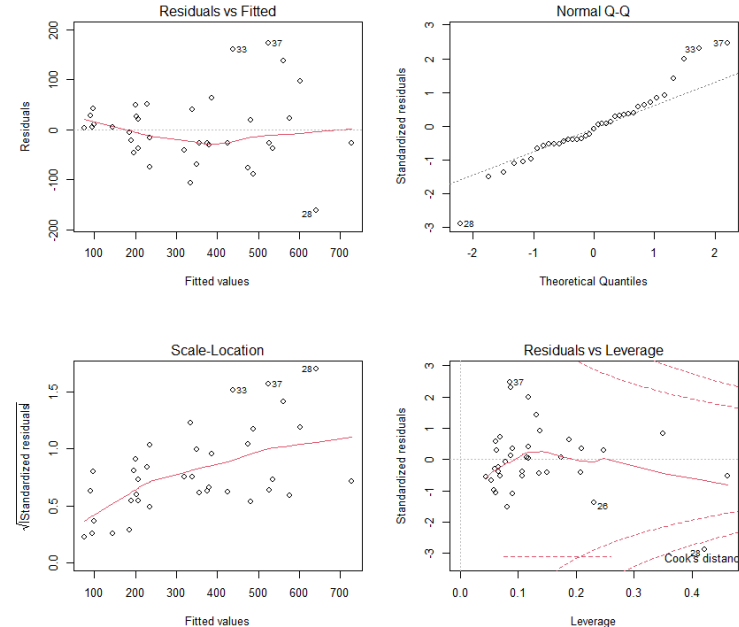
```
call: lm(formula = "precio de lanzamiento (USD) ~ rendimiento teórico (Tflops FP32) + 
    rendimiento en tests (FPS) + ancho de banda de la gráfica GB/s + 
    popularidad en videojuegos (steam)", data = MLibro)

Residuals:    Min       1Q   Median       3Q      Max 
-161.564   -36.924    -5.975    27.242   173.712 

Coefficients:
(Intercept)                0.02716 
rendimiento teórico (Tflops FP32)  -10.7631  4.6619  -2.309 
rendimiento en tests (FPS)         3.9385  0.7336  5.226 
ancho de banda de la gráfica GB/s    0.1347  0.2402  0.810 
popularidad en videojuegos (steam)  -615.1953  640.1946  -0.961 

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 73.72 on 32 degrees of freedom
Multiple R-squared:  0.8661,    Adjusted R-squared:  0.8494 
F-statistic: 51.75 on 4 and 32 DF,  p-value: 1.586e-13
```



Tras el análisis de este primer modelo observamos que las variables de “Popularidad en Videojuegos” y “Ancho de banda de la gráfica” no son significativas y por tanto tenemos que tenemos que eliminarlas del modelaje. Esto se debe a que que estos datos siguen un comportamiento que no se ajusta al modelo de comportamiento del “precio” ni del “rendimiento teórico”. Por otro lado, obtenemos un coeficiente de determinación corregido de 0,8494; por lo que podemos decir que nuestro modelo se ajusta con bastante precisión al precio real que podría tener una gráfica con determinadas características. Después de eliminar variables no significativas nos queda el modelo final:

```
call: lm(formula = "precio de lanzamiento (USD) ~ rendimiento teórico (Tflops FP32) + 
    rendimiento en tests (FPS)", data = MLibro)

Residuals:    Min       1Q   Median       3Q      Max 
-179.279   -32.440    -4.034    28.754   188.212 

Coefficients:
(Intercept)                0.02716 
rendimiento teórico (Tflops FP32)  -7.164  27.301  -0.262  0.7946 
rendimiento en tests (FPS)         4.543  4.330  -1.973  0.0507 

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 73.45 on 34 degrees of freedom
Multiple R-squared:  0.8588,    Adjusted R-squared:  0.8505 
F-statistic: 103.4 on 2 and 34 DF,  p-value: 3.533e-15
```

Lo que nos deja con la ecuación:

Precio= -7,164 - 8,543*Rendimiento teórico(Tflops) + 4,3*Rendimiento en tests(fps)

CONCLUSIONES

Podemos extraer varias conclusiones gracias a este estudio, la primera es que vemos que la correlación precio-popularidad no es lineal, como hemos visto en el modelo de regresión, en la misma tónica, el ancho de banda no es un factor determinante a la hora de decidir el precio de salida de las tarjetas, como también hemos podido observar en el modelo.

Otra conclusión clara es que el precio medio de las gráficas ha subido, en vez de bajar, como debería haber ocurrido en un mercado normal. Sin embargo esto es esperable ya que la situación en la actualidad es excepcional debido a la escasez de material de producción y a la alta demanda por el mercado de las criptomonedas.