

Precios de ordenadores portátiles

Alejandro Zubiri Funes

INTRODUCCIÓN Y OBJETIVOS

Se van a estudiar los precios de ordenadores portátiles en función del tamaño de la pantalla, su memoria RAM, su compañía, la compañía de su CPU, y la compañía de su GPU.

Se han obtenido los datos de Kaggle.

ESTUDIO DE CORRELACIÓN

Vamos a estudiar la correlación entre el precio y las diferentes variables cuantitativas. En concreto, correlación entre el precio y el tamaño de la pantalla y la RAM. Se ha estudiado el coeficiente de correlación, la covarianza, y la V de Cramer respecto a ambas:

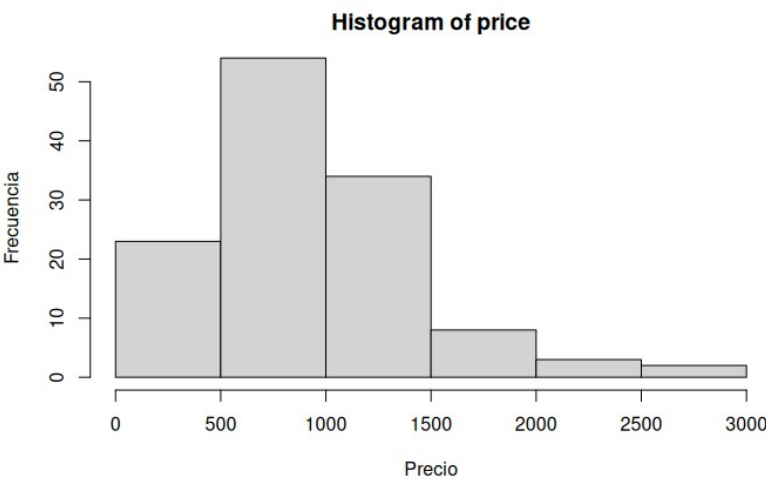
Precio respecto a:

- Ram
- Coeficiente de correlación: 0.6322
- Covarianza: 1087.835
- V de Cramer: 0.9323
- Tamaño de la pantalla
- Coeficiente de correlación: -0.1564
- Covarianza: -101.8424
- V de Cramer: 0.967

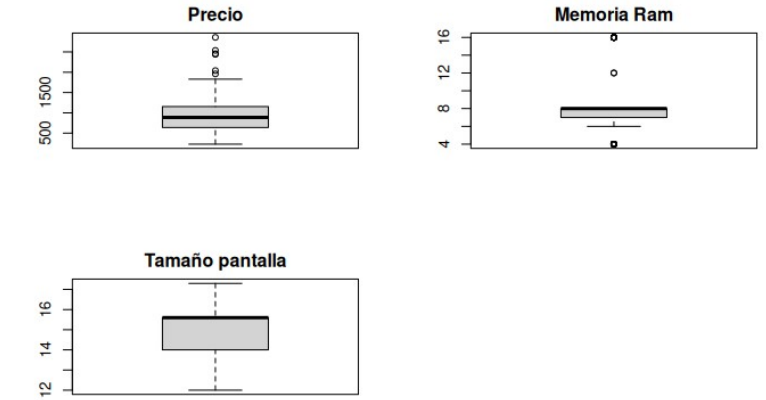
Con los siguientes resultados, podemos afirmar que hay una clara relación positiva del precio con la RAM, mientras que hay una relación negativa con el tamaño de la pantalla.

ESTUDIO DESCRIPTIVO

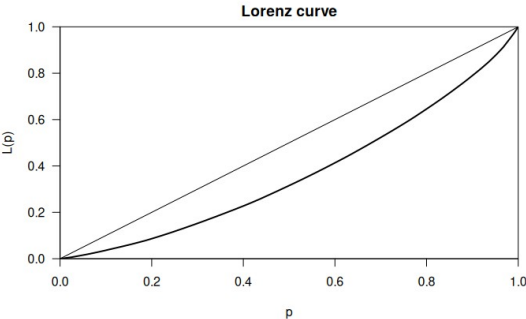
Empezamos estudiando la media, la desviación típica y la concentración. El histograma de los precios es el siguiente:



Donde se observa una clara cola y cierta asimetría positiva (media mayor a la mediana). La media es de 953.83, la desviación típica de 482.92, y la mediana es de 887.5. Además, el coeficiente de variación es de 0.5063, por lo que la media es poco representativa. El diagrama de caja de las variables cuantitativas es:



Se pueden observar numerosos datos atípicos, además de varias asimetrías tanto en los tamaños de las pantallas como en la memoria RAM. El índice de Gini de los precios es de 0.2748, lo que indica una baja concentración en los datos. La curva de Lorenz respectiva es



Además de esto, el máximo de los datos es de 2858 euros, y el mínimo de 229 euros. La curtosis es de 2.46, por lo que es leptocúrtica.

REGRESIÓN MÚLTIPLE

Si se realiza un primer modelo que incluya tanto variables cualitativas como cuantitativas, se encuentra que las variables no son independientes entre sí:

```
lm1 <- lm(formula = price ~ inch + ram + d1 + d2 + d3 + d4, data = dnm)
Residuals:
    Min       1Q   median       3Q      Max
-110.28  -107.38  -70.82   149.08  1827.08
Coefficients: (4 not defined because of singularities)
(Intercept)  513886  5417601  14408  92111
inch         -20.866   21.803   -1.94   0.554
ram          79.853    7.899    3.06   0.214
d1          146.172  100.404    4.863  3.3e-16 ***
d2          146.172  100.404    4.863  3.3e-16 ***
d3          146.172  100.404    4.863  3.3e-16 ***
d4          146.172  100.404    4.863  3.3e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 297.4 on 120 degrees of freedom
Multiple R-squared:  0.53,    Adjusted R-squared:  0.503
F-statistic: 88.12 on 5 and 120 DF, p-value: < 2.2e-16
```

En este caso, hemos incluido tanto la compañía, la CPU y la GPU. Después de eliminar esas variables, nos quedamos solo con el tamaño de la pantalla (inch) y la memoria RAM. El modelo que cumple las hipótesis y todas sus variables son significativas es el siguiente:

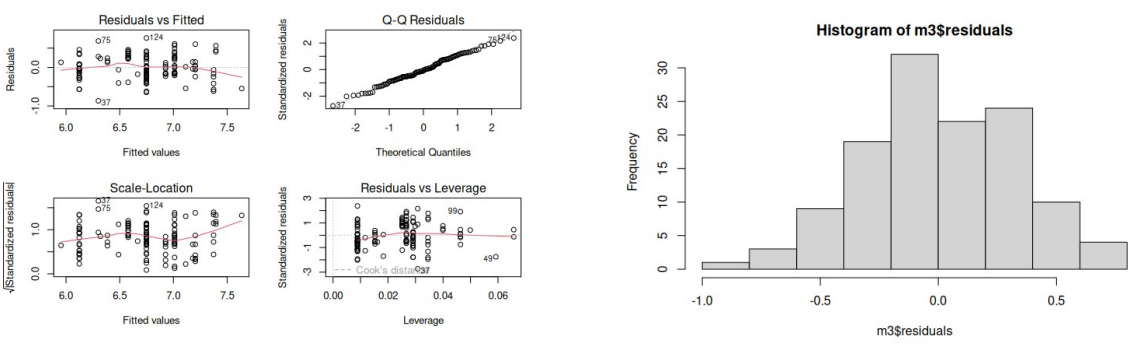
```
lm2 <- lm(formula = log(price) ~ log(inch) + log(ram), data = dnm)
Residuals:
    Min       1Q   median       3Q      Max
-0.86763  -0.21830  -0.02366   0.24733   0.76287
Coefficients:
(Intercept)  7.30665    0.87955  18.677  < 2e-16 ***
log(inch)   -1.64446    0.32448  15.898  5.46e-16 ***
log(ram)     0.90285    0.87052  12.752  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7327 on 121 degrees of freedom
Multiple R-squared:  0.584,    Adjusted R-squared:  0.563
F-statistic: 87.23 on 2 and 121 DF, p-value: < 2.2e-16
```

Siendo $\log(\text{precio}) = 9.39 - 1.644 \cdot \log(\text{tamaño}) + 0.902 \cdot \log(\text{ram})$

Debido a la característica transformación que sea ha aplicado al modelo, es complicado determinar como cambia el modelo cuando una de las variables aumenta a igualdad de las otras.

A continuación, observaremos que este modelo, además de contar solo con variables significativas, cumple con las hipótesis:



Gracias al primer gráfico, afirmamos que el modelo presenta linealidad, y con el segundo, que presenta normalidad, ya que los puntos se asemejan a una línea. Esto viene respaldado por el histograma de los residuos, que se asemeja a una distribución normal. El tercer gráfico nos indica que es homocedástico, ya que se observa una distribución aleatoria, además de que se aproxima a una línea. Como ahora sabemos que las variables son independientes, podemos concluir que nuestro modelo cumple con las hipótesis.

Con esto llegamos a que el modelo es capaz de explicar el 58.34% de la variabilidad de los datos.

CONCLUSIONES

Los precios de los ordenadores portátiles están muy dispersos y hay mucha variedad entre ellos, pero es claro que la mayoría de ellos cuentan con un precio comprendido entre 500 a 1500 euros, con portátiles de mayor precio posiblemente debido a marcas como Apple.

La dependencia de las compañías es algo predecible, teniendo en cuenta factores como la competencia de las marcas y patrocinios que las fuercen a utilizar marcas determinadas.

Respecto a las variables cuantitativas, es lógico que una mayor cantidad de memoria RAM aumente los precios, ya que proporcionan más potencia, pero una pendiente negativa respecto al tamaño de la pantalla se puede deber a razones de diseño o comodidad a la hora de transportar o usar el dispositivo.

En concreto, el portátil con el que se ha realizado el trabajo tiene 8 gigas de memoria RAM y una pantalla de 15.54 pulgadas. El modelo predice un precio de $\text{Precio} = e^{(9.39 - 1.644 \cdot \log(15.54) + 0.902 \cdot \log(8))} = 858.75$ euros

Costó alrededor de 900 euros, por lo que es una buena aproximación.