

Són igual de ràpids?

Comparació estadística entre bloquejadors d'anuncis.

Resum

- **Objectiu:** L'objectiu d'aquesta pràctica és comparar la velocitat de càrrega de diverses pàgines segons el bloquejador d'anuncis instal·lat. Volem comprovar si tots els bloquejadors d'anuncis són igual de ràpids.
- **Mètodes:** Per tal de capturar les dades, hem mesurat els temps de càrrega de 40 pàgines web amb publicitat. Primerament hem mesurat sense cap bloquejador d'anunci, i posteriorment amb cada un dels dos bloquejadors d'anuncis (AdBlock Plus, uBlock Origin). Quant al mètode d'anàlisi de dades hem fet servir un test aparellat entre el temps de càrrega de AdBlock Plus i uBlock Origin.
- **Resultats:** Els temps de càrrega recollits de les pàgines dels dos bloquejadors semblen ser millors que les mesures de les mateixes pàgines sense cap bloquejador. Dels dos bloquejadors fets servir, **un d'ells sembla ser més ràpid** alhora de carregar les pàgines. Al final de les proves de l'experiment, conclourem quin dels dos bloquejadors aporta una millor experiència d'usuari. **¿No se puede decir cuál?**
Hay que cuantificar la ventaja
- **Conclusió:** Després d'analitzar les mesures obtingudes hem arribat a la conclusió de que el bloquejador uBlock Origin fa que el temps de càrrega de les pàgines siguin més ràpids que amb el AdBlock Plus.

Introducció

A mesura que l'accés a internet ha anat entrant a les nostres vides, la publicitat també hi ha anat al darrere. Ha entrat tant en el món virtual, i en alguns casos ho ha fet de manera desagradable, que alguns programadors han desenvolupat bloquejadors d'anuncis que permeten a l'usuari amagar la publicitat que tenen les pàgines d'internet.

El fet, és que aquests bloquejadors d'anuncis no només amaguen la publicitat sinó que bloquejen moltes altres coses i fan que la càrrega de la majoria de pàgines web sigui més ràpida. Això comporta que l'experiència d'usuari alhora de consultar informació sigui millor.

Hem preparat la següent experiència per tal d'esbrinar si tots els bloquejadors que han anat apareixent en els darrers anys fan el mateix i si per tant tenen la mateixa efectivitat.

Mètode: Mesura de les dades

Abans de prendre les mesures dels temps de càrrega de les pàgines web es van definir les següents variables que ens permeten definir inequívocament l'eina amb la qual les mesures seran preses.

Vam prendre la decisió de fer servir els ordinadors de les aules de SO, ja que tots tenen les mateixes especificacions tècniques i podíem assegurar que la imatge del sistema operatiu seria sempre igual independentment del moment en que anéssim a prendre les mesures ja que la imatge es reestableix cada vegada que l'ordinador és reiniciat.

Els ordinadors de les aules de SO compten amb les següents característiques tècniques:

- Processador i7 3,4 GHz amb 4 nuclis i dos fils per nucli • 8 Gb de memòria RAM
- Arquitectura de 64bits

Quant al software que s'hi troba instal·lat tenim les següents característiques.

- Ubuntu 16.04 LTS
- Firefox 47.0 (deshabilitant el cache local)

Vam decidir fer servir el navegador firefox pels següents motius:

- és el navegador per defecte d'Ubuntu.
- a la seva botiga de complements hi havia més de un bloquejador d'anuncis compatible (<https://addons.mozilla.org/ca/firefox/search/?q=ad+block>).
- per defecte té instal·lada una eina que ens permet mesurar el temps de càrrega d'una pàgina (veure figura 1).
- permet configurar les opcions de memòria cau cosa que ens permet deshabilitar-la i minimitzar el seu efecte en les nostres dades.

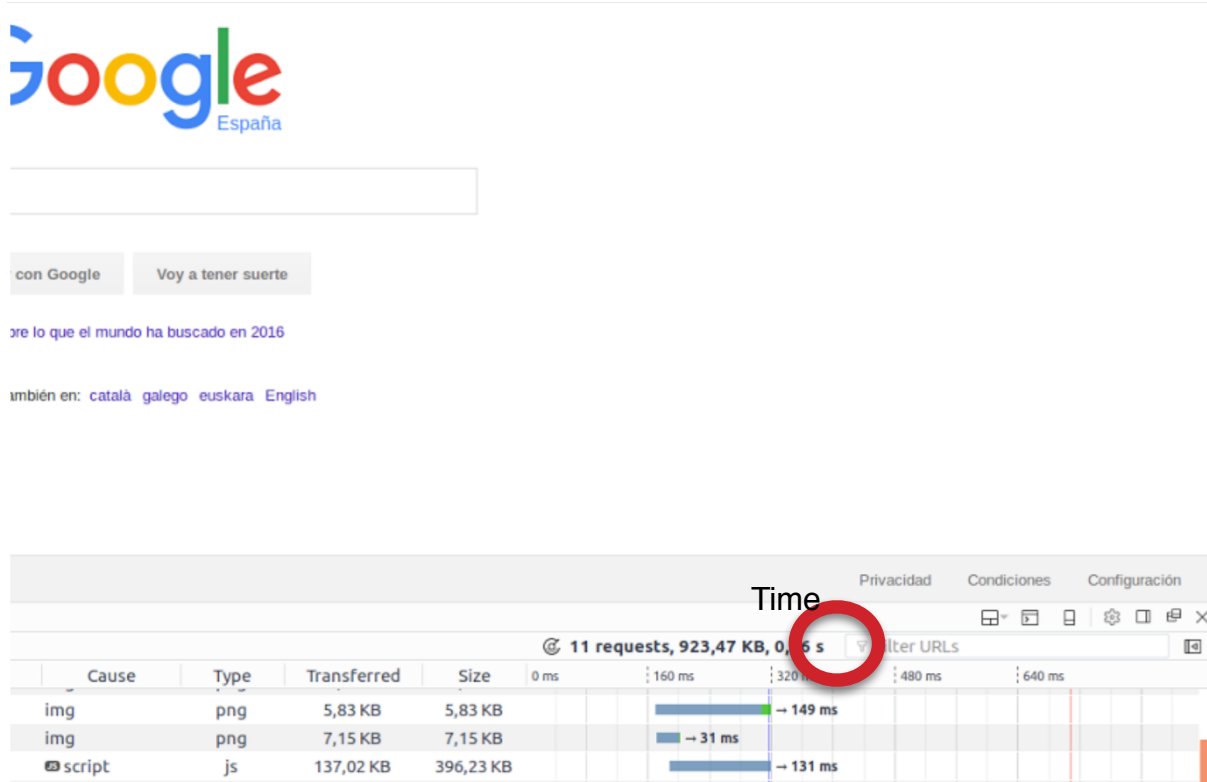


Figura 1: Eina nativa de Firefox per mesurar el temps de càrrega d'una pàgina.

Firefox per defecte no té cap bloquejador instal·lat així que després de cercar a la botiga de complementos hem decidit mesurar el temps de càrrega de les pàgines de les següents maneres:

- sense bloquejador d'anuncis instal·lat
- instal·lant l'AdBlock Plus (<https://addons.mozilla.org/ca/firefox/addon/adblock-plus/>)
- i instal·lant l'uBlock Origin (<https://addons.mozilla.org/ca/firefox/addon/ublock-origin/>).

Volem remarcar que tot i que sí que podem deshabilitar la memòria cau local gestionada per Firefox, no podem deshabilitar la memòria cau de tots els dispositius per els quals passen els paquets que demanem, per exemple quan des de la xarxa eduroam de la UPC s'intenta accedir a google es passa per 11 dispositius (Figura 2).

Això significa que el primer cop que es demana una pàgina carregarà més lentament que en successius intents, ja que en aquests algun dels dispositius ja tindrà carregada la pàgina a la memòria cau. Per tal de minimitzar aquest efecte vam decidir recarregar la pàgina 5 cops abans de no prendre una mesura definitiva.

```
gabriel@ubuntu-gabriel:~$ traceroute google.es
traceroute to google.es (172.217.19.131), 30 hops max, 60 byte packets
 1 gateway (10.192.1.1)  2.038 ms  2.291 ms  2.279 ms
 2 10.10.124.14 (10.10.124.14)  142.246 ms  142.564 ms  142.557 ms
 3 anella-upc.cesca.cat (84.88.18.17)  142.931 ms  142.919 ms  142.899 ms
 4 ANELLA-VAL1.AE2-454.uv.rt1.val.red.rediris.es (130.206.211.69)  143.309 ms
143.297 ms  143.278 ms
 5 UV.AE5.telmad.rt4.mad.red.rediris.es (130.206.245.90)  143.860 ms  143.833 m
s  143.815 ms
 6 google-router.red.rediris.es (130.206.255.2)  143.772 ms  361.563 ms  361.46
8 ms
 7 72.14.235.20 (72.14.235.20)  361.038 ms  72.14.235.18 (72.14.235.18)  204.616
ms  72.14.235.20 (72.14.235.20)  193.553 ms
 8 216.239.41.237 (216.239.41.237)  193.504 ms  193.486 ms  193.468 ms
 9 108.170.232.18 (108.170.232.18)  193.441 ms  209.85.255.138 (209.85.255.138)
193.426 ms  193.406 ms
10 66.249.94.79 (66.249.94.79)  148.847 ms  149.172 ms  150.543 ms
11 par03s12-in-f131.1e100.net (172.217.19.131)  193.918 ms  193.901 ms  70.598
ms
```

Figura 2: Dispositius per els quals passa un paquet quan va de un punt A a un punt B.

Tenint en compte les consideracions exposades anteriorment, vam definir el següent protocol per la captura de les dades.

1. Inici de la sessió.
2. Encendre ordinador i iniciar sessió.
3. Iniciem la captura de dades per una certa pàgina identificada per una url
4. Iniciar Firefox, instal·lació del bloquejador o no i configuració de la memòria cau.
 1. Mentre que el nombre d'accessos a url des del mateix terminal sigui més petit que 4
5. Recarreguem la pàgina un cop el símbol de càrrega es queda fixe.
6. Anotem el valor capturat pel Firefox en el moment que el símbol de càrrega es quedi fixe.
7. Si no hem mesurat el temps de càrrega per les opcions, tornem al pas 3 i canviem

Esquema 1: Protocol de captura de dades.

Mètode: Anàlisi de les dades

1. Càlcul de la diferència de càrrega de pàgines web amb diferents bloquejadors de publicitat

Definim A com el conjunt de temps de càrrega de les pàgines web amb el bloquejador AdBlock Plus i definim B com el conjunt de temps de càrrega de les pàgines web amb el bloquejador uBlock Origin.

Definim la variable D com la diferència de temps de càrrega de pàgines web entre un bloquejador i un altre (A-B).

Al ser mostres aparellades, cada subíndex de cada conjunt (A,B,D) corresponen a les mesures de la mateixa font, és a dir, la mateixa pàgina web.

Ex: $D_1 = A_1 - B_1$

2. Premisses convenients

1. Observar Normalitat en ambdues variables (A i B).
2. Càlcul bilateral per una certesa del 95%
3. Mostres aparellades aleatòries amb contingut bloquejant per ambdues variables.

3. Contrast d'hipòtesi a realitzar:

$$\begin{cases} H_0 : \mu_D = 0 \\ H_1 : \mu_D > 0 \end{cases} \quad \text{¿Por qué unilateral?}$$

4. Càlcul estadístic per la comparació:

Al ser dades aparellades realitzarem un test aparellat.

$$\hat{t} = \frac{\bar{D} - \mu_D}{S_D/\sqrt{n}} = (\text{si } \mu_D = 0) = \frac{\bar{D}}{S_D/\sqrt{n}} \sim t_{n-1}$$

- \bar{D} serà la mitjana de la variable diferencia D
- S_D serà la desviació típica de D
- μ_D serà el valor a contrastar: 0 per la igualtat de les 2 opcions

5. Obtenció del valor que delimita la Regió Crítica ($\alpha = 0.05$):

- Si $|T| > t_{n-1, 0.975}$ llavors es rebutja la hipòtesis nul·la.

6. Construcció de l'interval de confiança per la diferència:

$$IC \left[\bar{D} \mp t_{n-1, 0.975} \cdot \frac{S}{\sqrt{n}} \right]$$

Resultats

A partir dels càlculs realitzats amb el programa R, hem pogut determinar que A i B segueixen una Normal, tal i com es pot veure a la figura 3 i 4 tot i que la Normal es mostra esbiaixada. Un cop demostrat per a A i B, demostrem que la variable D també segueix una Normal, tal i com es pot veure a la figura 5 ja que els valors estan concentrats al centre de la recta.

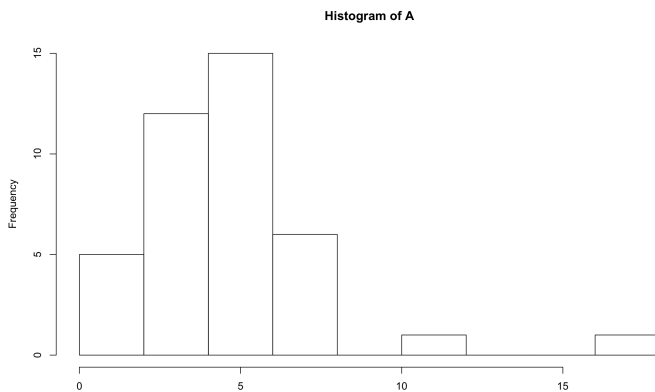


Figura 3: Histograma de A

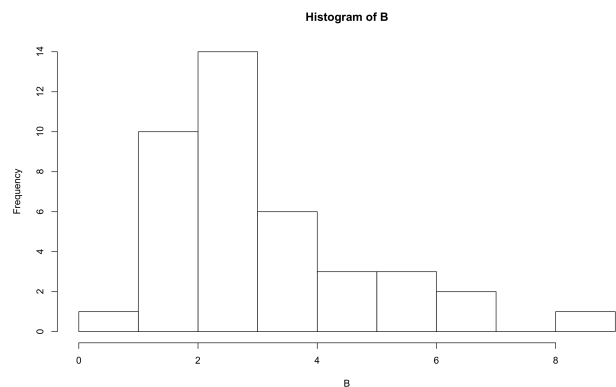


Figura 4: Histograma de B

Se echa de menos una descriptiva completa, medias, desviaciones, sobre todo para la diferencia.

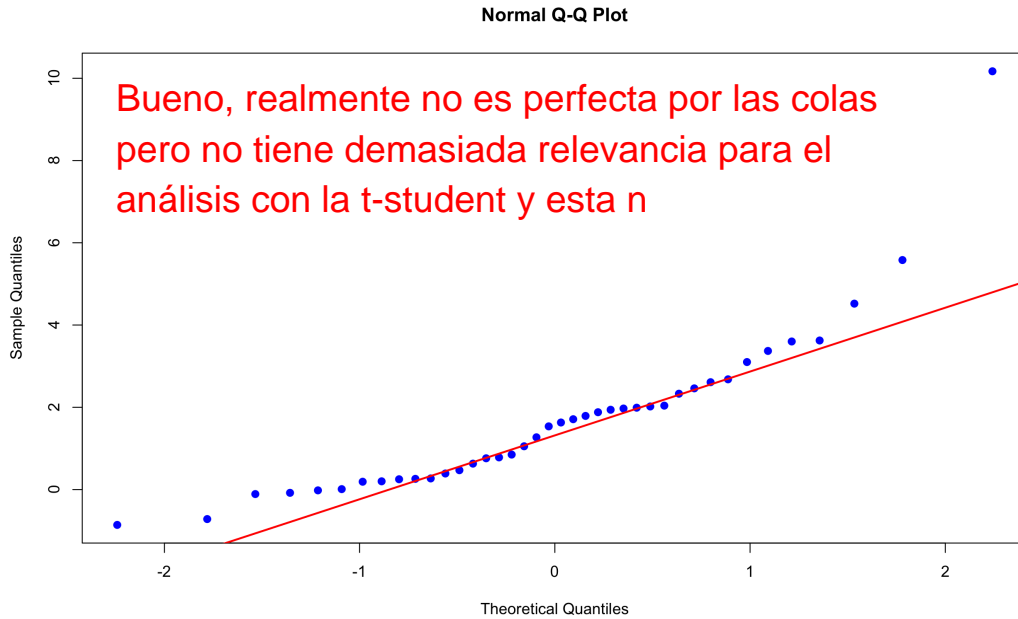


Figura 5: Gràfic Q-Q Normal

A continuació podem veure el contrast d'hipòtesi d'igualtat de variàncies que hem realitzat a través de R per tal de comparar les velocitats dels bloquejadors i mostrar que hi ha una diferència entre ambdós bloquejadors.

```
Paired t-test
```

```
data: D
```

```
t = 5.4415, df = 39, p-value = 3.085e-06
```

```
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
```

```
1.070378 2.336922
```

```
sample estimates:
```

```
mean of the differences
```

```
1.70365
```

y esto es bilateral

Esquema 2: Resultat d'aplicar el t-test a R sobre D. [`>t.test(A,U,paired=TRUE,var.equal=TRUE)`]

A través del t-test hem pogut obtenir el p-valor amb un resultat de $3.085 \cdot 10^{-6}$. Amb aquest valor podem refutar la Hipòtesi Nul·la ja que es compleix que:

$$3.085 \cdot 10^{-6} < 0.05$$

Aquesta condició és més que suficient per tal de refutar la Hipòtesi Nul·la, de manera que queda demostrat que els dos bloquejadors no són iguals de ràpids.

A partir dels resultats anteriors, podem afirmar amb un IC de [1.070378, 2.336922] que el bloquejador uBlock Origin és més ràpid alhora de carregar pàgines respecte a l'Adblock Plus perquè recordem que D es la diferència entre A i B, i l'interval es positiu.

$$IC \left[\bar{D} \mp t_{n-1, 0.975} \cdot \frac{S}{\sqrt{n}} \right] = [1.070378, 2.336922]$$

Analitzant la Figura 6 podem concloure dues coses:

- El bloquejador Adblock Plus acostuma a trigar més alhora de carregar les mateixes pàgines que uBlock Origin, el qual triga menys.
- Les 40 mostres agafades mostren certa aleatorietat.

A partir de la Figura 6, també podem veure que la majoria de resultats estan per sobre de 0, la qual cosa vol dir que la majoria de diferències és positiva i per tant favorable a la idea de que uBlock Origin és més ràpid que Adblock Plus.

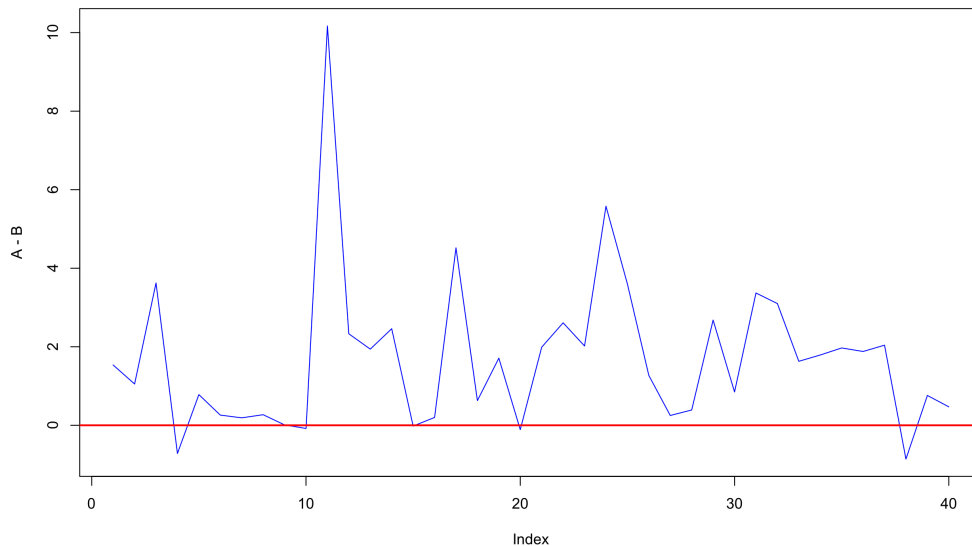


Figura 6: Gràfic de línies de D

Discussió

A partir de l'anàlisi de les dades hem conclòs que:

- Els dos bloquejadors no són iguals.
- uBlock Origin sembla obtenir millors temps de càrrega (més ràpid) que AdBlock Plus.

Per acabar de consolidar l'interval de confiança, ens hauria agradat obtenir més mesures. No obstant les dades obtingudes ens permeten treure conclusions.

No hay limitaciones realmente. ¿No se os ha ocurrido nada más?
Adicionalmente, podríais haber incluido los tiempos sin bloqueador,
aunque fuera sin análisis estadístico.

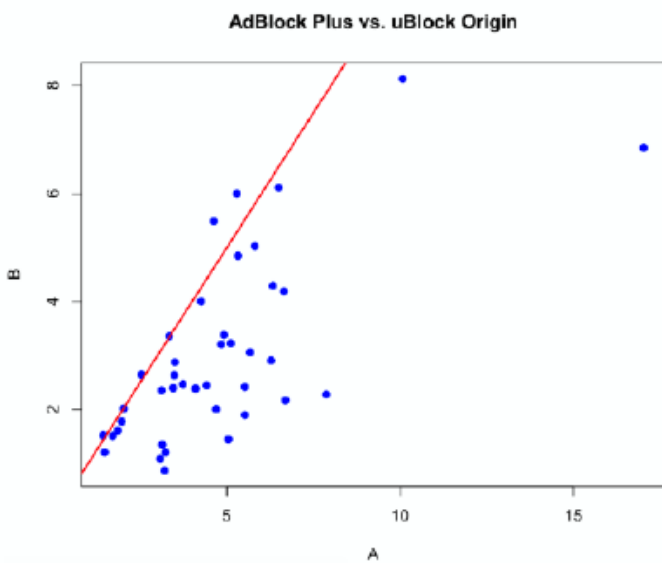


Figura 6: AdBlock Plus vs. uBlock Origin

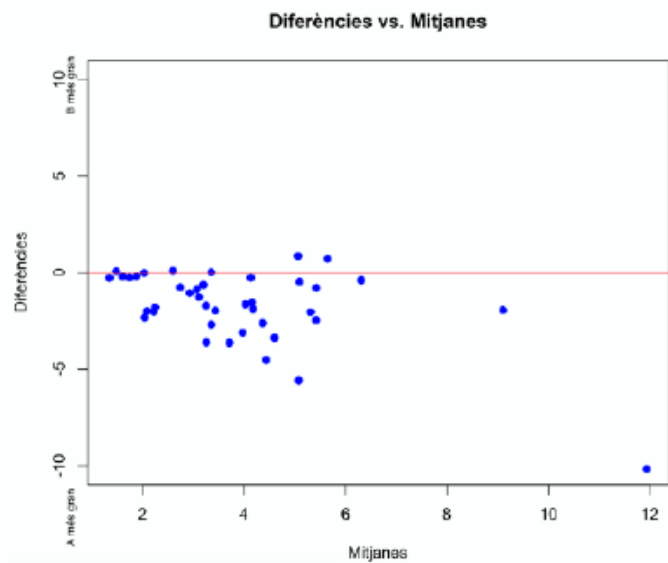


Figura 7: Diferències vs. Mitjanes

Fe de erratas: el informe definitivo de los estudiantes sí incluía estos dos gráficos imprescindibles, el que relaciona dos variables apareadas (a la izquierda), y el de Bland-Altman (a la derecha). Sobre todo en este se aprecia que cuando el tiempo de carga es pequeño no hay apenas diferencia entre los dos bloqueadores.