Estudi sobre els navegadors webs més utilitzats per a visitar una determinada pàgina web.

Aida Albalate, Ferran Arau, Oscar Iglesias

Resum

Objectiu: Es vol relacionar el nombre de visites que rep al dia una pàgina web amb el navegador que s'ha emprat per accedir-hi.

Mètodes: Hem agafat una mostra de 52 dies, des del 3 de febrer fins el 15 d'abril (sense tenir en compte els caps de setmana) utilitzant l'eina de Google Analytics per extreure les dades que necessitàvem. Voliem saber el nombre de visites per dia i els navegadors utilitzats en aquestes visites.

Resultats: Les dades obtingudes mostren que s'han fet en promig menys visites utilitzant Internet Explorer. Les mitjanes ens han donat 4.019 amb IE i 8.241 per a la resta de navegadors.

Discussió: Hem arribat a la conclusió de que en aquesta pàgina web no cal emprar tècniques especials de maquetació per a Internet Explorer. Encara que també hem pensat que s'hauria de fer un estudi sense comptabilitzar les visites fetes per la pròpia empresa, per tal d'obtenir resultats més fidedignes.

Introducció

En el món de la maquetació web existeix la gran problemàtica que hi ha navegadors que no segueixen els estàndards. En concret el navegador Internet Explorer és, a priori, un dels navegadors més utilitzats, i també el que menys segueix els estàndards. Per tant, per optimitzar la maquetació de la web de la empresa en la que basem el nostre estudi seria útil conèixer el navegador més utilitzat per al públic d'aquesta web per emprar unes tècniques o unes altres.

Objectiu

L'objectiu d'aquest estudi és relacionar el nombre de visites de la pàgina web de la empresa amb el navegador que s'ha emprat per accedir-hi, per saber si s'ha d'optimitzar la maquetació web per al navegador Internet Explorer o no.

Material y mètodes

Tenim accés a un compte de Google Analytics[1] de la pàgina web d'una petita empresa. Aquest aplicatiu registra informació de control del tràfic web.
S'ha estudiat el nombre de visites per dia, i les visites per explorador web.
Diferenciarem els navegadors en dos grups, internet explorer i els altres (Firefox, Safari, Chrome, Opera).

Hem agafat una mostra de 52 dies, des del 3 de febrer fins el 15 d'abril, sense tenir en compte els caps de setmana, ja que la pàgina web de la qual fem l'estudi té un públic que majoritàriament són empreses, per tant les dades de cap de setmana no són rellevants.
Les mostres consten del nombre de visites per navegador i dia.
La variable resposta en aquest estudi és el nombre de visites i la variable explicativa és el navegador.

Anàlisi estadística

La descriptiva de les variables numèriques es realitza a través de les mitjanes i les desviacions estàndards (DE).
A l'anàlisi principal es van seguir els següents passos:

Prova de significació

En el nostre experiment tenim mostres aparellades, donat que cada dia (unitat que utilitzem per agafar les mostres) dóna lloc a dues mesures. Per tant, per fer l'estudi necessitarem definir una nova variable anomenada D, que serà la diferència entre la variable noIE (visites fetes amb altres navegadors) i la variable IE (visites fetes amb internet explorer).

Nombre de mostres (dies) n = 52
Total visites: 662
noIE: 445 (67.22%)
IE: 217 (32.78%)

Dades de la variable diferència D:
Mitja: 4.222222
Variància: 20.93082

[1] Per a més informació: http://www.google.com/intl/es/analytics/

Mitjançant el software R hem obtingut les dades necessàries per desenvolupar la Prova de Significació de l'estudi.

I. Hipòtesi:

H_0: $\mu_D = 0$

H_1: $\mu_D \neq 0$

(prova bilateral)

II. Estadístic i distribución sota H_0:

$t = (D - 0) / (S_D / \sqrt{n})$ t_{n-1}

$t \rightarrow t_{n-1}$

t_{52-1}

III. Premises

- Les mostres són **aparellades** com hem comentat a l'inici de la prova de significació.
- La variable D (diferència) segueix una distribució normal com es pot apreciar a la gràfica 2.

IV. Càlculs

Càlcul del valor de l'estadístic

$t = (4.2222 - 0) / (4.575021/\sqrt{52}) = 6.7818$

Càlcul del p-valor

p-value = 1.012e-08

Interval de confiança amb el 95% de confiança:

[2.973482 - 5.470963]

V. Decisió

Rebutgem la hipòtesi si $t > t_{52-1}$.

L'estadístic t era igual a 6.7818, per tant, està fora del interval de confiança proporcionat per la T-Student amb una probabilitat del 95%. Això implica que el p-valor de la mostra serà inferior a α (0.05 > 1.012e-08). Aleshores no hi ha prou evidència per acceptar H0.

Resultats

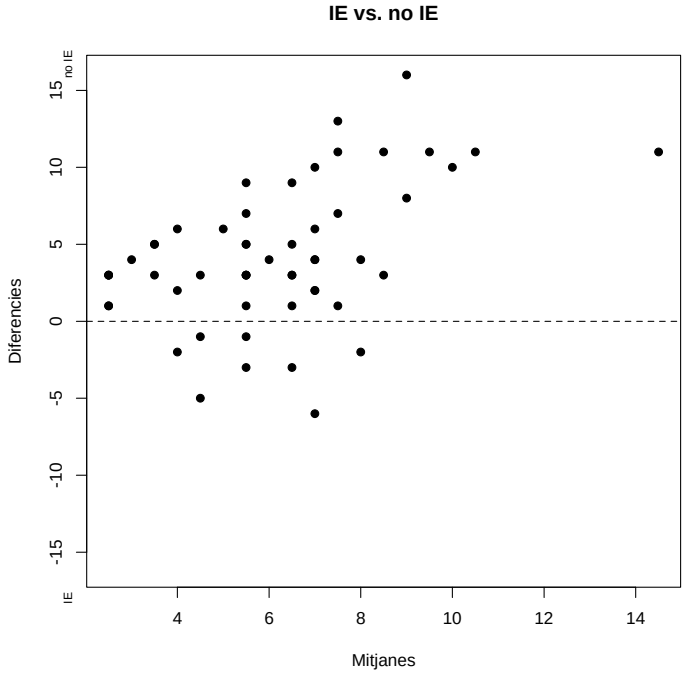
Es desitja saber (PS) si la variable NoIE (visites fetes amb altres navegadors) millora a la variable IE (visites fetes amb internet explorer).

Descriptiva

La taula 1 mostra la comparació del nombre de visites a la página amb el navegador I.Explorer, enfront els altres navegadors.

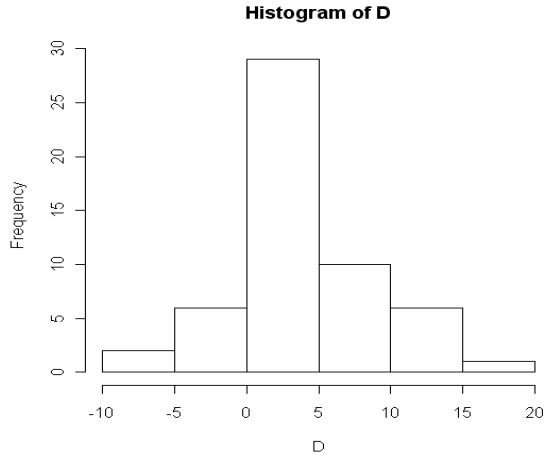
Taula 1: Descriptiva de les variables

	Global	IE	No IE
Mitjana	12.26	4.019	8.241
Desviació estandar	4.654822	2.359171	3.966535
1er QU	9.00	2.00	5.00
3rd QU	14.75	5.00	10.00
Min	5.00	1.00	2.00
Max	29.00	10.00	20.00
mediana	12.50	4.00	8.00

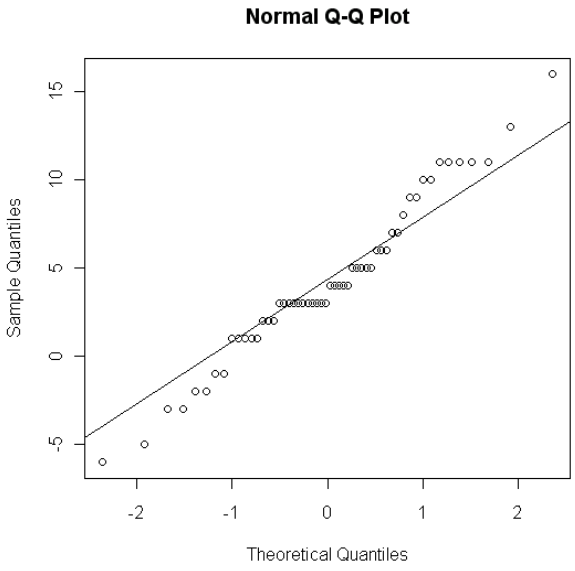


Grafica 1: Diferència de mitjanes aparellades (amb efecte aditiu)

A la gràfica 1 es veu clarament que hi ha una major concentració de punts a la meitat superior, per tant hi ha més visites utilitzant noIE. Tal i com s'indica a l'eix vertical, si els punts s'acosten a la part inferior, significa que s'utilitza més internet explorer i si s'apropen a la part superior, tot el contrari.



Gràfica 2: Histograma de la variable diferència



Gràfica 3: qqnorm de la variable diferència

A la gràfica 2 podem veure que la variable diferència (D) segueix una distribució Normal, per tant, es compleix la premissa de Normalitat que ha de seguir una prova de mostres aparellades.

En canvi a la gràfica 3 no es compleix gaire bé la normalitat. Això és degut a que s'utilitzen nombres enters.

Discussió

Conclusió principal:

Els resultats obtinguts són diferents dels que esperàvem a priori, ja que, en bon principi era de suposar que internet explorer tindria molta més afluència que la resta, res més lluny de la realitat. Però alhora de fer les proves, tots els altres acaben tenint més pes com es demostra a l'estudi.
Si ens fixem a la gràfica 1, veiem clarament que quasi tots els punts estan a la meitat superior, indicant que hi ha més visites fetes amb altres navegadors que amb internet explorer.
Podem concloure amb aquests resultats que el públic que visita la pàgina web del nostre estudi majoritàriament utilitza altres navegadors diferents a internet explorer, amb la qual cosa, els maquetadors de la pàgina web no han d'utilitzar tècniques específiques per internet explorer.

Limitacions de l'estudi:
Ens va semblar interesant contemplar el temps de visita que estava cada usuari per tenir en compte també si els usuaris que visitaven la pàgina web durant més estona utilitzaven internet explorer o no, però Google Analytics no recull aquesta informació.
A més a més, les dades recollides també inclouen les visites de la pròpia empresa, que no accedeix amb el navegador internet explorer, amb la qual cosa els resultats poden estar lleugerament afavorits cap a la conclusió que hem arribat.

Futurs estudis:
Recomanem activar una opció a l'eina de Google Analytics per no comptabilitzar les visites fetes per la pròpia empresa per a que en futurs estudis no quedi reflectida aquesta decantació.

ANNEX I. Scripts en R:

```
# Carreguem les dades per al R
total <-c(
11,20,14,15,14,14,18,5,17,19,7,18,11,15,11,8,29,7,17,11,5,15,7,16,11,8,13,13,5,5,12,6,11,9,15,
11,10,11,13,5,9,21,16,13,13,11,9,8,14,14,13,14,11,14)
nie <- c(
7,15,4,13,10,9,13,4,14,15,5,17,10,11,8,7,20,6,10,5,3,8,6,7,4,3,7,9,4,4,8,5,9,4,14,8,8,7,11,3,6,16,
10,8,5,7,2,5,8,12,8,8,6,9)
ie <- c(
4,5,10,2,4,5,5,1,3,4,2,1,1,4,3,1,9,1,7,6,2,7,1,9,7,5,6,4,1,1,4,1,2,5,1,3,2,4,2,2,3,5,6,5,8,4,7,3,6,2,5
,6,5,5)

# Descriptiva de les dades
summary(total)
summary(ie)
summary(nie)
sd(total)
sd(ie)
sd(nie)

#Script per a generar les gràfiques descriptives
BlandAltman <- function(x,y,tit) {
        Bmean <- (x+y)/2
        Bdif <- y-x
        ymax <- max(abs(Bdif))
        plot(Bmean ,Bdif ,
                ylim=c(-ymax,ymax),
                xlab="Mitjanes",
                ylab="Diferencies",
                main=tit,pch=19)
        abline(h=0,lty=2)
        mtext("no IE",2,line= 0.5,
                at=1.1*ymax,adj=1,cex=0.7)
        mtext("IE",2,line= 0.5,
                at=-1.1*ymax,adj=1,cex=0.7)
}
```

```
plot(ie,nie,main="IE vs. no IE")
BlandAltman(ie,nie,"IE vs. no IE") -> qqnorm(Y-X)

t.test(nie,ie,paired=TRUE,var.equal=True)
      Paired t-test
          t = 6.7818, df = 53, p-value = 1.012e-08
          alternative hypothesis: true difference in means is not equal to 0
          95 percent confidence interval:
           2.973482 5.470963
          sample estimates:
          mean of the differences
                      4.222222

D = nie-ie
hist(D)
```