COMPARACIÓ DE CERCADORS DE VOLS ONLINE



Resum:

Objectiu:

Comparant dos dels principals cercadors de vols online volem veure quin dels dos és més ràpid en qualsevol cerca. 

Mètodes:

S'ha fet un recull de dades en situacions independents per a diversificar el mostreig, però sempre conservant la igualtat de condicions (Hardware, hora i dia, trajectes cercats...) entre ambdós cercadors.



Introducció:

Actualment es pot veure com cada cop més s'està popularitzant l'ús d'Internet com a eina per a dur a terme moltes de les activitats quotidianes tals com transaccions bancàries, la compra, etc. Entre aquests serveis destaquen els cercadors de vols online. Igual com en altres casos a Internet, es tracta d'un servei massificat, i podem trobar infinitud de cercadors diferents. D'entre aquest ampli ventall de possibilitats seria interessant averiguar trobar algun que destaqui per la seva eficiència.



Objectiu:

Havent sel·leccionat dos dels cercadors més populars i coneguts, atrapalo.com i edreams.es, es vol averiguar quin dels dos és més ràpid a l'hora de realitzar les cerques, contemplant el temps entre que se sol·licita la cerca i es completa la càrrega de la pàgina de resultats.



Materials i mètodes:

Per tal de cronometrar els temps hem fet servir un plugin cronòmetre de Google Chrome (page load time) executat sobre Windows XP (en ordinadors de la FIB) amb característiques: Intel Core 2 Duo E8400 a 3.0GHz i 2GB de memòria RAM.
Hem obtingut les mostres a partir de diferents tipus de trajectes per tal d'aconseguir la millor representació de cada cercador (viatges nacional -> nacional, nacional -> internacional, internacional -> nacional, internacional -> internacional). També hem variat les dates dels trajectes però sempre fent servir les mateixes dates per a un mateix trajecte en els dos cercadors.

Hem establert una variable resposta corresponent al temps de càrrega dels cercadors.



Procés d'estudi

1. Prova de significació:

És unilateral, volem comprovar que la mitjana d'un dels cercadors es significativament menor que la de l'altra.

H_0: $\mu_a = \mu_e$

H_1: $\mu_a \; != \mu_e$

2. Premisses:

- Les V.A. es corresponen a una distribució normal.

- La prova és unilateral.

- Tenim mostres aleatòries.

3. Estadístic i distribució:

μ	$H_0 : \mu_1 = \mu_2$	$\hat{t} = \frac{(\bar{y}_1 - \bar{y}_2)}{S\sqrt{1/n_1 + 1/n_2}}$ $S^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$	$Y_1, Y_2 \rightarrow N$ $\sigma_1 = \sigma_2$ m.a.s indep.	$\hat{t} \rightarrow t_{n1+n2-2}$	Rebutjar si $\lvert \hat{t} \rvert > t_{n1+n2-2,\ 1-\alpha/2}$ $\lvert \hat{t} \rvert > t_{n1+n2-2,0'975}$ amb α=5%

Fig. 1. Hipòtesis, premisses i distribució de l'estudi.

Resultats:

Hem obtingut els següents resultats mitjançant el programa R:

```
mean(atrapalo)
7.22825
median(atrapalo)
3.19
```

```
mean(edreams)
19.16825
median(edreams)
17.25
```

Podem observar que la mitjana de la variable *atrapalo* és significativament menor que la de *edreams*. Si tenim en compte la diferència entre les mitjanes i les medianes de cadascuna de les variables, podem deduir l'existència d'"outlayers" significatius, especialment en el cas d'*atrapalo*.

```
min(atrapalo)          min(edreams)
1.4                    3.11
max(atrapalo)          max(edreams)
34.4                   57.3
```

Observant aquestes dades, podem remarcar la disparitat de resultats entre les dues variables, un cop més a favor de la variable *atrapalo*.

Per a poder contemplar aquestes conclusions de forma gràfica, hem obtingut el boxplot d'ambdues variables:

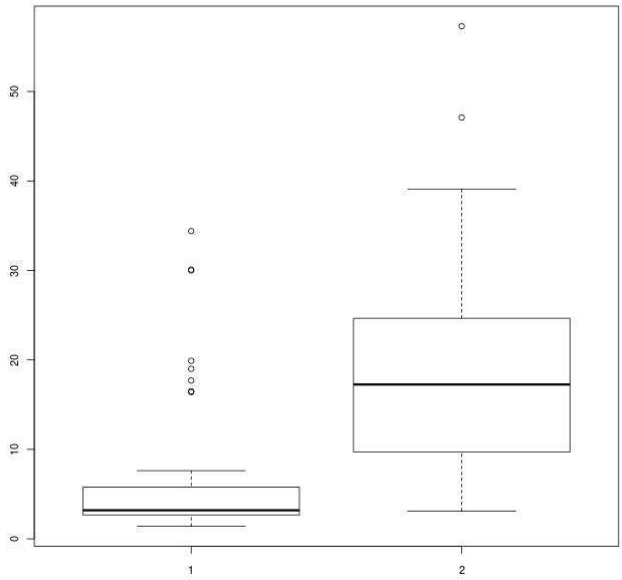


Fig. 2. Boxplots de les variables *atrapalo* i *edreams*, respectivament.

Podem observar clarament una gran diferència entre les dades recollides.

Seguidament, comprovem la premissa de normalitat de les variables mitjançant `qqnorm` de R:

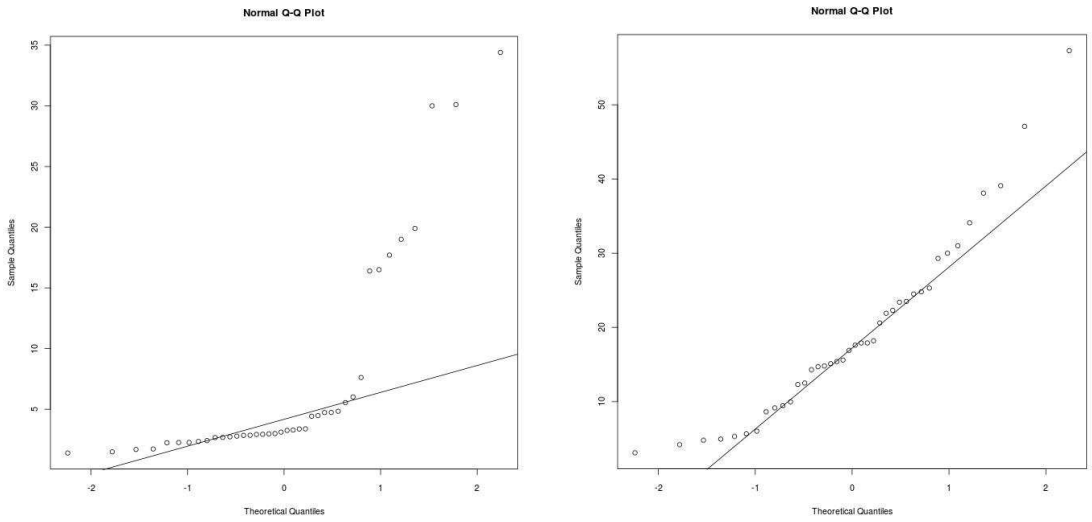


Fig. 3. QQplots de les variables *atrapalo* i *edreams* respectivament.

Tot i la abundant presència d'"outlayers" en el cas de la variable *atrapalo*, podem afirmar que en ambdós casos les variables compleixen normalitat.

Finalment, havent comprovat totes les premisses, i de nou mitjançant R, durem a terme la prova d'hipòtesi, amb la comanda `t.test`, obtenint els resultats següents:

```
data:  atrapalo and edreams
t = -5.0616, df = 70.414, p-value = 3.189e-06
alternative hypothesis: true difference in means is not
equal to 0
95 percent confidence interval:
 -16.644295  -7.235705
sample estimates:
mean of x mean of y
 7.22825  19.16825
```

Discussió

La conclusió del nostre estudi aboga a favor del cercador *atrapalo*, ja que, tot i que ens podríem trobar en un moment puntual amb una cerca que trigui més de lo normal ("outlayers"), hem pogut comprovar que en general, les cerques de qualsevol tipus són més ràpides que si usem el cercador d'*edreams*. A més, tenint en compte els pitjors casos en ambdós cercadors, com podem veure si observem els temps màxims recollits de cadascun, la conclusió segueix favorable per *atrapalo*.