

Estudi de la ocupació de les aules de la FIB

Arnau Canyadell Miquel

Joan Marcè Igual

Daniel Ferro González

2 de juny de 2015

Índex

1	Resum	1
2	Introducció	1
2.1	Objectiu	1
3	Metodologia	1
3.1	Recollida de dades	1
3.2	Variables d'estudi	1
3.3	Contrast d'hipòtesis	2
3.4	Premisses	2
3.5	Estadístic utilitzat	2
4	Resultats	2
4.1	Descripció de les dades	2
4.2	Comparació de la disponibilitat	3
5	Discussió	4
5.1	Conclusió principal	4

1 Resum

L'objectiu d'aquest treball és comparar la disponibilitat d'ordinadors entre aules de la FIB on hi ha classe i aules on no n'hi ha i comprovar si és cert que hi ha més ordinadors lliures en proporció a les aules on hi ha classe. La metodologia seguida és la recollida de dades mitjançant un servidor i accedint a l'API de la FIB i la comparació de les mitjanes de les proporcions d'ordinadors no ocupats en aules on hi ha classe i aules on no partint de la hipòtesi nul·la que aquestes mitjanes són iguals. El resultat, comparant el p-valor és que la hipòtesi nul·la és falsa perquè la proporció mitjana d'ordinadors lliures en aules on no hi ha classe és menor. A partir d'aquest resultat arribem a la conclusió que la hipòtesi inicial era falsa i ho atribuïm al fet que hi ha moltes hores en què hi ha poca gent a la universitat i hi ha molt poques classes.

2 Introducció

A la FIB és molt comú anar en busca d'una aula d'informàtica per tal de fer un treball o altre. A nosaltres ens va picar la curiositat per saber quina era la ocupació mitjana d'aquestes aules i a quines hores i quins dies estaven menys ocupades per tal de poder trobar-ne una lliure amb facilitat. Aquestes dades només serviran per aquest quadrimestre ja que cada quadrimestre les assignatures que es fan a la FIB canvien el seu nombre d'alumnes matriculats i les aules on es realitzen.

2.1 Objectiu

El nostre treball consisteix en analitzar la ocupació de les aules d'ordinadors de la *Facultat d'Informàtica de Barcelona*. Volem saber quant ocupades estan les aules i com es distribueix aquesta ocupació al llarg dia.

La nostra hipòtesi és que *en les aules on hi ha classe la disponibilitat és superior que en les que hi ha classe*. Per demostrar-ho, compararem el percentatge de disponibilitat d'ordinadors en les aules on es fa classe i les aules on no se n'hi fa.

3 Metodologia

3.1 Recollida de dades

Per a fer la recollida de dades, hem programat un servidor perquè es connecti a l'API de la FIB cada minut des del dimarts 5 de maig fins el divendres 29 del mateix mes i des de les 8 del matí fins a les 9 del vespre (horari d'obertura de la facultat). L'API del racó ens proporciona la informació següent:

- L'horari de classes de cada aula.
- El nombre d'ordinadors lliures per cada aula en l'instant en què es demana.

El programa utilitzat es pot compilar i executar des del següent enllaç:

<https://github.com/jmigual/peFIB>

3.2 Variables d'estudi

Per tal de fer-nos una idea de com és la disponibilitat de les aules, hem decidit treballar amb les variables següents:

X = proporció d'ordinadors lliures en les aules on *hi ha classe*.

Y = proporció d'ordinadors lliures en les aules on *no hi ha classe*.

3.3 Contrast d'hipòtesis

$$H_0 : \mu_x = \mu_y$$

$$H_1 : \mu_x \neq \mu_y$$

3.4 Premisses

1. Les variables aleatòries X i Y es poden aproximar a una normal degut a la seva grandària (11934 i 11951, respectivament).
2. Les diverses mostres preses de X i Y són consecutives i per tant no independents. Tanmateix, sí que podem suposar que al cap d'una hora d'haver pres una mostra, la nova mostra ja serà independent de l'anterior (suposem que hi ha prou moviment a les aules). Així doncs, si agaféssim només una de cada 60 mostres (1 cada hora), tindríem poblacions mostrals de, aproximadament, 200 individus, que són prou grans com per poder dir que X i Y s'aproximen a una normal.

3.5 Estadístic utilitzat

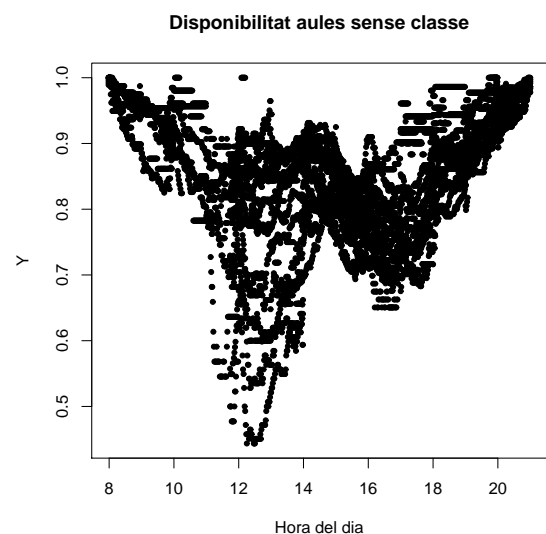
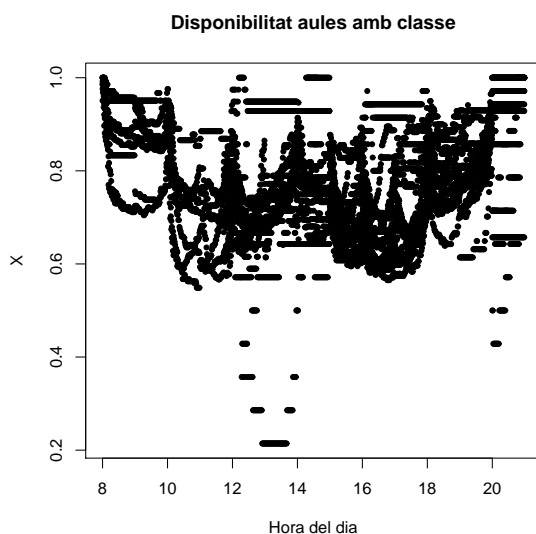
$$\hat{z} = \frac{\bar{x} - \bar{y}}{\sqrt{s_x^2/n_x + s_y^2/n_y}}$$

$$\hat{z} \sim N(0, 1)$$

Rebutjar si $|\hat{z}| > z_{1-\alpha/2}$ amb $\alpha = 5\%$

4 Resultats

4.1 Descripció de les dades



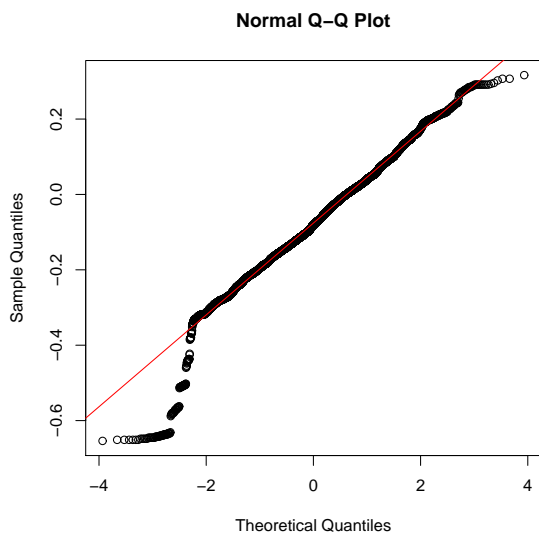


Figura 1

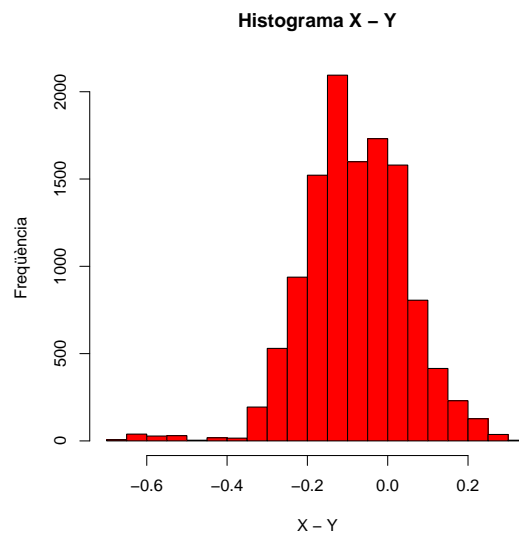


Figura 2

Els dos primers gràfics són els reculls de dades per les aules amb classe i les aules sense classe. Cada punt és la proporció d'ordinadors lliures en una hora en concret.

En les premisses hem suposat que l'estadístic \hat{z} es podia aproximar a una normal. A les figures 3 i 4 podem veure que les variables X i Y s'assemblen lleugerament al gràfic d'una distribució normal. La figura 1 mostra la desviació de $X - Y$ respecte a una normal.

Observant aquesta i la figura 2 es pot veure que la diferència entre X i Y (que és proporcional a \hat{z}) segueix una normal amb una petita anormalitat en els valors més petits.

Les dades completes es poden trobar en un full de càlcul seguint el següent enllaç:

<https://docs.google.com/spreadsheets/d/1e1GJjdaar26Jyu9gmVH1C1y-p9IX4dDBx0xu2t6Q0hw/pubhtml>

4.2 Comparació de la disponibilitat

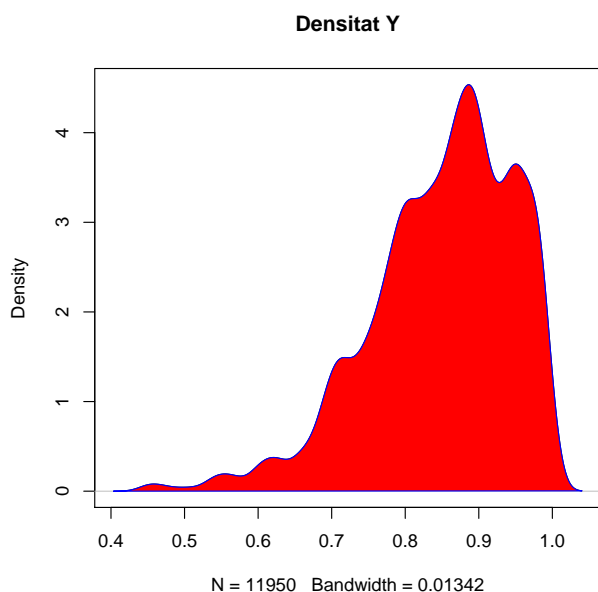


Figura 3

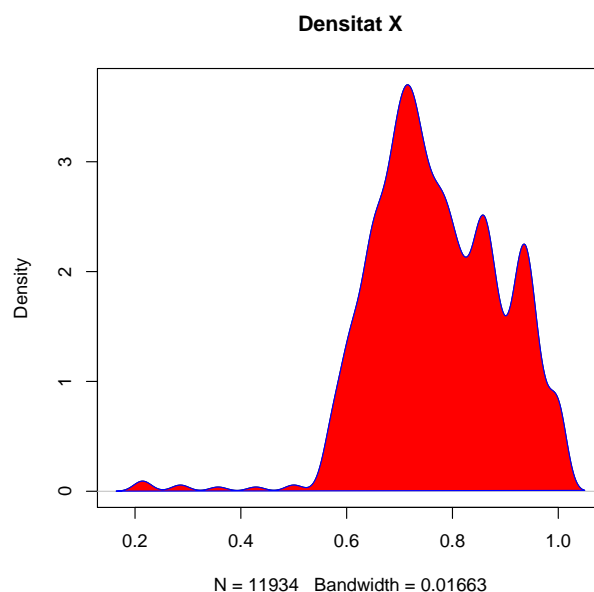


Figura 4

Analitzant les dades obtingudes hem arribat als següents valors:

$$\bar{x} = 0.77083, s_x^2 = 0.01459139, n_x = 11934$$

$$\bar{y} = 0.8503347, s_y^2 = 0.009507904, n_y = 11950$$

I per tant:

$$\hat{z} = -55.96263$$

El valor de \hat{z} és extremament petit. De fet, calculant el p-valor amb R per aquesta \hat{z} el resultat és 0. Per tant, queda clar que s'ha de rebutjar la hipòtesi nul·la. És més, degut a que \hat{z} ha resultat ser negativa, podem afirmar que $\mu_Y > \mu_X$.

5 Discussió

El comportament de les variables X o Y pot dependre de molts factors. Tot seguit n'enumerem uns quants:

Hora: els hàbits rutinaris dels estudiants i els horaris de classe fan que en certes hores hi hagi més gent utilitzant els ordinadors.

Ocupació de les aules per fer classe: com més aules estiguin reservades per fer-hi classe, menys ordinadors lliures hi haurà pels estudiants que no tenen classe i en volen fer servir un.

Exàmens: els exàmens amb ordinador, a part d'ocupar moltes aules, poden provocar que hi hagi molts estudiants repassant a l'últim moment i que l'ocupació en les hores abans sigui anormalment alta.

Això fa que les dades, tot i ser immenses, no ens proporcionin una informació tan completa.

5.1 Conclusió principal

Hem obtingut que la mitjana de disponibilitat de les aules amb classe és inferior a les mitjana de les aules sense classe. Això contradiu la nostra hipòtesis inicial del treball ja que creïem que les aules amb classe tindrien una disponibilitat superior a les sense classe.

Creiem que una bona explicació a aquest fenomen és el fet que hi ha moltes hores, sobretot al vespre, en què hi ha molt poques classes i també hi ha pocs alumnes utilitzant les aules lliures, i això fa baixar la mitjana. Nosaltres ens basàvem en la impressió que tenim a les hores punta, en què és difícil trobar un ordinador lliure i en canvi hi ha classes on no s'omplen tots els ordinadors. Podria ser interessant analitzar la disponibilitat en aquestes hores en concret.