Taxes de fumadors a la FIB i Telecomunicacions

Pau Castells, Jordi Cebrián, Manel Sellés

24 de maig del 2011

Resum

Crec que l'objectiu real és saber si existeixen diferències entre la proporció de fumadors a la FIB i a la ETSETB. Aquí esteu dient que compararem, però em sembla que el fet de comparar no és un objectiu, sinó que formaría part de la metodologia

Objectiu: Determinar quina es la població fumadora a la Facultat d'Informàtica de Barcelona (FIB) i comparar-la amb la població fumadora d'una altra facultat.

Inidicar que és la ETSETB

Potser millor dir: entre el 2 i el 6 de maig

Mètodes: Entre el 2 de maig i el 6 de maig de 2011 es van recollir les dades de 53 alumnes de la FIB i de 53 alumnes de l'Escola Tècnica Superior d'Enginyeria de Telecomunicacions de Barcelona (ETSETB) , referents al hàbit de fumar. Es va examinar la relació entre fumadors i no fumadors mitjançant un model binomial.

A l'hàbit

També heu utilitzat un model que segueix una distribució Chi-Quadrat, i caldria afegir-ho. Podrieu indicar que vàreu fer servir també una prova de significació i contrast d'hipòtesis per decidir si les proporcions semblen iguals o no per arrodonir-ho

Resultats: El p-valor obtingut ha estat de 0.07503, i el estadístic 3.1694

l'estadístic

Conclusió: No podem demostrar que la probabilitat de ser fumador variï segons si s'estudia a la FIB o a l'ETSETB.

El terme "demostrar" pot sembla un pèl massa fort en l'àmbit de la inferencia estadística. Potser seria més idoni dir alguna cosa d'aquest estil: No podem creure que la probabilitat de ser o no ser fumador depengui de la facultat a on s'estudiï

Fer servir la primera persona del plural

Introducció

Potser quedaría millor dir: "en què es trobin"

Els hàbits de la gent poden variar depenent del lloc, sexe o del moment on estiguin situades. Un d'aquests hàbits és el de fumar. L'estudi de la proporció de població fumadora es un dels més realitzats en la societat actual.

Objectius

Potser quedaría millor dir: "de manera notòria, en proporció, el nombre"…

L'objectiu d'aquesta pràctica és determinar si es pot diferenciar proporcionalment i de manera notable el nombre de fumadors a la FIB amb el nombre de fumadors a l'ETSETB.

Que futures generacions d'estudiants de Probabilitat i Estadística puguin utilitzar aquestes dades recollides per comparar-les amb les que hagin obtingut i poder comparar-les amb les que hagin recollit.

Elimina perquè l'expressió es repetix

Material i mètodes

Accent

Es va proposar recollir la informació sobre l'habit de fumar de diferents estudiants de la FIB en diferents llocs, dies i hores. Aquells mateixos dies també vam preguntar a estudiants de ETSETB seguint la mateixa metodologia. Entre el 2 de maig i el 6 de maig de 2011 es van recollir 53 dades de cada població.

Potser millor dir: entre el 2 i el 6 de maig

El terme población crec que està mal emprat aquí. Crec que seria millor dir que de cada facultat vau obtenir una mostra de 53 observacions

Anàlisi estadística

La descriptiva de les variables numèriques es realitza a traves de les dades recollides i les distribucions marginals.

Em sembla que us falta indicar explícitament quina és la vostra variable resposta i quina l'explicativa

Cal mencionar que vau fer servir enquestes per a la recollida de dades

Fer servir la primera persona del plural

Resultats

Descriptiva

La taula 1 descriu les dades obtingudes en les enquestes classificant-les, en fumadors no fumadors i en poblacions.

Freqüència observada: f_{ij}	Facultat d'Informàtica	Escola de Telecomunicacions	Total
Fumen	17	26	43
No Fumen	36	27	63
Total	53	53	N=106

Taula 1: Dades recollides.

A partir d'aquestes dades podem veure que el percentatge de fumadors a ETSETB es una mica superior al de la FIB.

Accent

Anàlisi principal → Crec que us falta indicar les premisses (i si puguessiu validar-les millor), indicar explicitament quina és la vostra hipótesis nul·la i alternativa, i indicar en cada moment si esteu realitzantr una PS o bé un CH. També heu d'indicar si el vostre punt de vista és bilateral o unilateral (i per quina banda en aquest últim cas)

Calculem les distribucions marginals

Freqüència esperada: e_{ij}	Facultat d'Informàtica	Escola de Telecomunicacions	Total
Fumen	21,5	21,5	43
No Fumen	31,5	31,5	63
Total	53	53	N=106

Taula 2: Distribucions marginals

Calculem l'estadístic a mà, utilitzant la següent fórmula. → Quedaria millor, em sembla, si indiquessiu el nom que rep aquest estadístic

$$\hat{X}^2 = \sum_{\forall ij} \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

Pel càlcul de l'estadístic seguirem la següent distribució: → Crec que seria mes idoni si doneu el nom de la distribució i expliciteu el nombre de graus de llibertat (em refereixo a dir-ho per escrit)

X^2_1

- Càlcul del valor de l'estadístic

$(f_{ij}-e_{ij})^2/e_{ij}$	Facultat d'Informàtica	Escola de Telecomunicacions	Total
Fumen	0,941860	0,941860	
No Fumen	0,642860	0,642860	
Total			3,16943493

El valor de l'estadístic que hem calculat manualment ha donat 3.16943493

Ara utilitzant el programa R realitzarem els càlculs seguint una distribució normal i una de Pearson.
Els valors obtinguts seguint la distribució normal han sigut:

```
> prop.test(c(17, 26),c(53,53), correct=FALSE)

        2-sample test for equality of proportions without continuity
        correction

data:  c(17, 26) out of c(53, 53)
X-squared = 3.1694, df = 1, p-value = 0.07503
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.35394447  0.01432182
sample estimates:
   prop 1    prop 2
0.3207547 0.4905660
```

Figura 1: Càlculs mitjançant Normal → Crec que s'hauria de ser més precís i dir el següent (o similar): Càlculs mitjançant una prova amb l'estadísitc z que sota H0 té una distribució Normal

El valor obtingut(Figura 1) ha sigut el mateix que hem obtingut de calcular-lo manualment. ↘ Calculant-lo

Amb un p-valor = 0.07503 i un interval de confiança del 95% [-0.35394447, 0.01432182]

"una prova de Pearson" em sembla que seria més adient ↗

Ara comprovarem si amb una Chi-quadrada de Pearson obtenim els mateixos valors.

```
 FIB Telecos
n  36      27
f  17      26

> taula = read.table("clipboard",header=TRUE)
> chisq.test(taula, correct=FALSE)

        Pearson's Chi-squared test

data:  taula
X-squared = 3.1694, df = 1, p-value = 0.07503
```

Figura 2: Càlculs mitjançant Pearson → Pot quedar més acurat si dieui: "Càlculs mitjançant una prova amb l'estadístic de Pearson (o Chi Quadrat)"

Podem veure com els valors obtinguts son els mateixos que amb la distribució normal en la figura 2.

Discussió

Trobo que és una conclusió correcta, però si s'utilitzés un llenguatge més tècnic quedaría encara millor. Proposo una discussió del següent estil: "el p-valor és prou versemblant, per tant, no tenim suficient evidència com per creure'ns que les proporcions de fumadors entre les dues facultats siguin diferents. En altres paraules, no hem trobat evidències prou significatives com per dubtar de la homogeneïtat en ambdues mostres (independencia entre facultat i ser o no ser fumador)

Els valors obtinguts no han sigut els esperats, el p-valor és prou versemblant com per no poder demostrar que la probabilitat de ser fumador pot variar segons si estudies a la FIB o a l'ETSETB.

Amb un p-valor = 0.07503 i un interval de confiança del 95% [-0.35394447, 0.01432182] no podem acceptar la hipòtesis nul·la

REBUTJAR

A l'hora de

Hem estat limitats alhora de poder escollir els estudiants completament a l'atzar. No podíem utilitzar una llista amb tots els estudiants de la FIB i una altra amb els de ETSETB i fer-ne una selecció completament aleatòria. Això pot haver deteriorat la qualitat del nostre estudi.

Accent

Un altre aspecte a tenir en compte es el nombre d'enquestats, segurament amb un nombre major l'estudi hagués sortit de més qualitat.

Per a la recerca futura d'aquest mateix tema s'hauria de tenir en compte aquests aspectes mencionats recentment.

Cal afegir les posibles extrapolacions que pugueu observar en el vostre estudi

En general, el contingut crec que és l'adequat, tot i que es podría aprofundir una mica més i aportar més material (com gràfics en l'anàlisis principal o descriptiva). Sou molt concissos i us centreu en l'estudi. Tot el contingut està ben estructurat, seguint un ordre lògic i pautat. No s'ha omès pràcticament cap apartat dels que calia presentar. El format del treball segueix el model presentat en anteriors exemples. I expresseu de manera força clara els conceptes explicats a teoría i sense divagar o fer-se repetitiu.