

DIPLOMATURA D'ESTADÍSTICA

**Projecte Final de Carrera.
Diplomatura d'Estadística.**

**Títol: Impacte dels Grans Equipaments
Comercials en el territori.**

**Alumnes: Josep Ramon Marsal Mora
Carlos Pellicer Alapont**

Director de Projecte: Alexandre Riba Civil

UNIVERSITAT POLITÈCNICA DE CATALUNYA
Biblioteca



1400458109



**Facultat de Matemàtiques
i Estadística**

UNIVERSITAT POLITÈCNICA DE CATALUNYA



PROJECTE FINAL DE CARRERA:

**IMPACTE DELS
GRANS EQUIPAMENTS COMERCIALS
EN EL TERRITORI**

DIRECTOR DE PROJECTE: ALEXANDRE RIBA CIVIL

**ALUMNES: JOSEP RAMON MARSAL MORA
CARLOS PELLICER ALAPONT**



AGRAÏMENTS:

Volem donar les gràcies, en primer lloc, al nostre director de projecte, Alexandre Riba Civil, per la seva atenció, ajuda i consells en l'elaboració del present projecte.

També volem donar les gràcies a les nostres respectives famílies i parelles per la seva comprensió i suport incondicional en aquest important moment.

Finalment, i no menys important, cada un dels membres de la parella que fem aquest projecte agraeix a l'altre el temps sacrificat i esforç dedicat en aquest objectiu comú.

Moltes gràcies a tots.

ÍNDEX

	<u>Pàgines</u>
Capítol 1. Justificació, Objectiu, Abast i Antecedents	1
1.1 Justificació	2
1.2 Objectius	2
1.3 Abast	3
1.4 Antecedents	3
Capítol 2. Introducció Teòrica	4
2.1 Model de regressió lineal	5
2.2 Model de regressió logística	6
2.2.1 Definició de regressió logística	6
2.2.2 Transformació Logit	8
2.2.3 Model Simple de regressió logística	9
2.2.4 Model Múltiple de regressió logística	9
2.2.5 Estimació dels paràmetres del model de regressió logística	10
2.2.6 Comparació model de regressió lineal vs. model de regressió logística	12
2.2.7 Deviança del model	13
2.2.8 Altres usos de la regressió logística	15
Capítol 3. Base de Dades	16
3.1 Descripció gràfica de la base de dades	17
3.2 Composició de la base de dades	18
Capítol 4. Anàlisi exploratòria de les dades	22
4.1 Anàlisi Univariant	23
4.1.1 Variables de Població	23
4.1.2 Variables dels Grans Equipaments	24
4.2 Anàlisi Bivariant	25
4.2.1 Dades de municipis/districtes	25
4.2.2 Dades dels hipermercats i dels Grans Equipaments Comercials	26
4.2.3 Dades del comportament del consum	29
Capítol 5. Metodologia	31

	<u>Pàgines</u>
Capítol 6. Procés de modelat	34
6.1 Com ha de ser el model?	35
6.2 Primera part de la modelització	36
6.3 Segona part de la modelització	37
6.4 Tercera part de la modelització	41
Capítol 7. Resultats i Interpretació	43
Capítol 8. Conclusions	48
8.1 Conclusions del model	49
8.2 Exemple	50
Bibliografia	52
Annexos	
A.1 Matriu de gràfics bivariants i Correlacions per variables de població	A-1
A.2 Matriu de gràfics bivariants i Correlacions per variables dels hipermercats i GEC's	A-2
A.3 Matriu de gràfics bivariants i Correlacions per variables de consum	A-3
A.4 Matriu de gràfics bivariants i Correlacions per variables del model final	A-4
A.5 Models finals i sortides de Minitab	A-5
A.5.1 Primer Model	A-5
A.5.2 Segon Model	A-7
A.5.3 Model Definitiu	A-9
A.6 Sortides de Minitab pel Model Definitiu segons Zona Metropolitana	A-11
A.6.1 Zona Metropolitana de València	A-12
A.6.2 Zona Metropolitana de Madrid	A-13
A.6.3 Zona Metropolitana de Barcelona	A-14
A.7 Taula de la Base de Dades	A-15

Capítol 1

**Justificació,
Objectiu,
Abast i
Antecedents.**

1.1 JUSTIFICACIÓ

Les autoritats competents en matèria de comerç estan preocupades per la implantació massiva dels Grans Equipaments Comenrcials (GEC) ja que afecten a l'economia del petit comerç.

Els GEC's ofereixen una gran gamma de serveis i preus compatibles amb les necessitats dels consumidors e inaccessibles pel petit i mitjà comerç.

Per tant, s'han concieniat en conèixer i eval.luar l'impacte d'aquests tipus de comerç al voltant de les seves zones d'influència.

Aquest estudi va estar motivat per les conselleries de comerç de Catalunya, Madrid i València.

1.2 OBJECTIUS

Es vol modelar l'impacte dels grans equipaments comercials (GEC's) i en concret els hipermercats d'aquests sobre el territori.

El què interessa a les autoritats competents, és la proporció de diners que es deixen de gastar els clients d'un municipi/ districte, en concret en el petit i mitjà comerç, i es gasten en l'hipermercat d'un GEC determinat. Però és difícil, preguntar a un consumidor els diners gastats l'hipermercat, per tant tenim una restricció. Una altra restricció serà que hi ha hipermercats que no autoritzen preguntar als seus clients els diners que s'han gastat.

Per tant si el que es vol és modelar la proporció de facturació d'un determinat hipermercat que asumeix un determinat municipi/ districte, el més fàcil serà modelar la proporció de gent.

L'objectiu del treball, és doncs, modelar la proporció de clients d'un municipi/districte que acudeixen a l'hipermercat d'un GEC en funció de la zona d'influència.

1.3 ABAST DE LA BASE DE DADES

Per modelar el comportament de la gent que acudeix a l'hipermercat d'un GEC, s'utilitzarà una base de dades ja recollida. El contingut de la base de dades s'ha tret de unes enquestes fetes anteriorment.

Les dades que hi ha en aquesta base van ser recollides en tres zones metropolitanes, en concret: Barcelona, Madrid i València i les seves àrees d'influència. Per tant, no s'ha d'oblidar que les conclusions a les que s'arribarà estan limitades a aquestes àrees metropolitanes i als seus voltants, i que possiblement no es podrà extrapolar la informació obtinguda a la totalitat de l'Estat, ja que no disposarem de més dades per comparar-ho.

1.4 ANTECEDENTS

L'únic precedent conegut per nosaltres és un projecte registrat com:

"Resolució de dos problemes reals mitjançant el modelatge estadístic".

Autor: Ramón Peñas Cambray.

Director: Alexandre Riba Civil.

Concretament, fem referència al primer del dos problemes tractats en aquest projecte.

Capítol 2

Introducció teòrica

2. INTRODUCCIÓ TEÒRICA

La regressió és una metodologia estadística que es basa en l'ús de la relació entre dues o més variables mesurables i tenint una variable que pot ser explicada a partir de l'altre/s

De manera formal, descomposarem la variable resposta (Y), en una part explicada a partir de les altres variables (X 's) i una part no explicable (ε) que és l'error.

Aquesta relació la podem representar com:

$$Y = E(Y/X) + \varepsilon \quad \text{o també} \quad Y = f(x_i) + \varepsilon$$

2.1 MODEL DE REGRESSIÓ LINEAL:

El model de regressió més conegut és el de regressió lineal; aquest model és adequat per aquelles relacions entre unes variables independents quantitatives i/o binàries i una variable dependent quantitativa que es poden descriure amb línies rectes.

Si ho expressem en forma d'equació:

$$Y = f(x_1, x_2, \dots, x_n) + \varepsilon_i \Rightarrow Y = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_{p-1} x_{i,p-1} + \varepsilon_i$$

On: Y : variable resposta

β_j : paràmetres del model

x_j : variables explicatives

ε_i : errors aleatoris que segueixen una distribució $N(0, \sigma^2)$ i són independents i estan idènticament distribuïts.

El que s'intenta es trobar la millor manera d'ajustar el nostre model lineal al model teòric; per fer això hem de trobar els coeficients β_j (del nostre model) que ajustin millor a les b_j (del model teòric).

Podem definir els residus com la part no explicada pel model i, per tant, són, per cada observació, la diferència entre el nostre model i el model teòric.

Així, el més habitual per ajustar el model experimental al model teòric és mitjançant el criteri dels mínims quadrats, que consisteix en minimitzar la suma dels quadrats de la part no explicable (residus)

2.2 MODEL DE REGRESSIÓ LOGÍSTICA:

Per l'elaboració de la introducció teòrica referida a la Regressió Logística hem consultat bàsicament dos llibres:

- "Análisis Multivariante en Ciencias de la Salud. Modelos de Regresión" U.D 13.
Editorial Signo S.A.
- Capítol 14 del llibre "Applied Linear Statistical Models"
Editorial Irwin.

Per veure més detalls de la bibliografia, s'han d'anar al Capítol 9 (Bibliografia) del present projecte.

2.2.1. Definició de la regressió logística:

Es un model específicament dissenyat per estudiar relacions entre una variable resposta categòrica i una o més variables explicatives que poden ser numèriques o categòriques.

Així, considerant la variable resposta Y_i com una variable que segueix una distribució de Bernoulli i que té una distribució de probabilitat tal com:

Y_i	Probabilitat
1	$P(Y_i=1) = \pi_i$
0	$P(Y_i=0) = 1 - \pi_i$

es pot assumir que π_i és la probabilitat de que $Y_i = 1$, y que $1 - \pi_i$ és la probabilitat de que $Y_i = 0$.

- **Hipòtesis bàsiques en el model de regressió logística:**

Quan la variable resposta és una variable indicadora tenim diversos problemes; com a exemple, utilitzant un model simple de regressió lineal podem mostrar-ne tres d'ells.

Pot ser, el més important de tots és el tercer punt i és el que ens indica que pel nostre model és correcte utilitzar el model de regressió logística en lloc del model de regressió lineal.

a. Els termes d'error (ε_i) no es distribueixen de manera normal.

Per una variable resposta binària (0,1), cada terme d'error $\varepsilon_i = y_i - (\beta_0 + \beta_1 x_i)$ pot prendre només dos valors:

$$\text{Quan } y_i = 1 \Rightarrow \varepsilon_i = 1 - \beta_0 - \beta_1 x_i$$

$$\text{Quan } y_i = 0 \Rightarrow \varepsilon_i = -\beta_0 - \beta_1 x_i$$

Clarament, el model de regressió lineal, que assumeix que ε_i està distribuït de manera normal, no és l'apropiat.

b. La variança dels termes d'error (ε_i) no és constant.

Un altre problema amb el termes d'error ε_i es que no tenen igual variança quan la resposta variable és una variable indicadora; pel model de regressió logística podem obtenir el valor de $\sigma^2\{y_i\}$

$$\sigma^2\{y_i\} = \pi_i (1 - \pi_i) = E\{y_i\} (1 - E\{y_i\})$$

i, com que la variança d' ε_i , és la mateixa que y_i , ja que $\varepsilon_i = y_i - \pi_i$ (i π_i és una constant) és té:

$$\sigma^2\{y_i\} = \pi_i (1 - \pi_i) = E\{y_i\} (1 - E\{y_i\})$$

c. Restriccions de la Funció Resposta.

Com que la funció resposta representa probabilitats quan la variable resultat és 0 o 1 (variable indicadora) les respostes mitjanes estan restringides així:

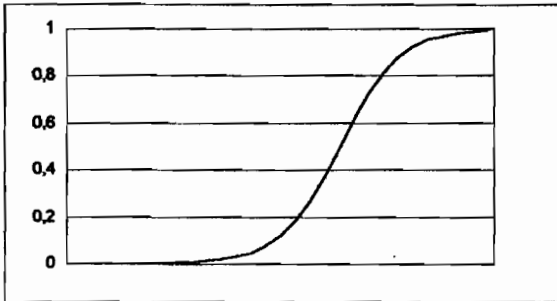
$$0 \leq E\{y\} = \pi \leq 1$$

La resposta d'una funció lineal, per exemple, pot caure fora dels límits de les restriccions dintre del rang de la variable explicativa en l'àmbit del model.

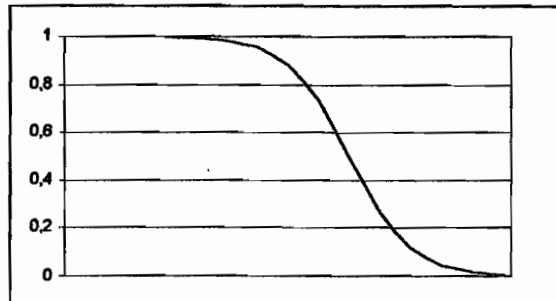
Per tant, amb una variable resposta binària, si modelem la proporció de clients d'una població que acudeixen a l'hipermercat amb un model de regressió lineal, podem obtenir valors de $E\{y/x\}$ negatius o superiors a 1 i això, és clar, que no és correcte.

Les consideracions teòriques i experimentals suggereixen que quan la variable resposta és binària, la forma de la funció resposta és freqüentment curvilínia.

Així, per exemple, en el cas més simple, amb una sola variable explicativa:



(figura 2.2.1)



(figura 2.2.2)

Les funcions resposta dibuixades en les figures 2.2.1 i 2.2.2 s'anomenen funcions resposta logístiques i són de la forma:

$$E\{y\} = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} \quad (\text{equació 2.2.1}),$$

o de la seva forma equivalent: $E\{y\} = [1 + \exp(-\beta_0 - \beta_1 x)]^{-1}$

2.2.2. Transformació Logit:

Una de les propietats més interessants de la funció resposta logística és que pot ser linealitzada fàcilment, de fet, aquesta propietat és molt important pel nostre projecte.

Si anomenem π a $E\{y\}$, podem fer la següent transformació:

$$\pi' = \ln\left(\frac{\pi}{1 - \pi}\right) \quad (\text{equació 2.2.2})$$

I l'equació 2.2.1 ens queda transformada en: $\pi' = \beta_0 + \beta_1 x$

Aquesta transformació s'anomena transformació logit de la probabilitat π .

La funció resposta transformada $\pi' = \beta_0 + \beta_1 x$ s'anomena resposta logit mitja.

2.2.3. Model Simple de Regressió Logística:

Si la variable resposta és binària, prenent els valors 1 i 0 amb probabilitats π i $1 - \pi$, respectivament, Y és una variable aleatòria de Bernoulli on $E\{Y\} = \pi$.

Podem escriure el model simple de regressió logística en la forma usual:

$$Y_i = E\{Y_i\} + \varepsilon_i$$

i, com que Y_i són variables aleatòries de Bernoulli amb valors esperats $E\{Y_i\} = \pi_i$, es pot presentar el model simple de regressió logística com:

$$E\{Y_i\} = \pi_i = \frac{\exp(\beta_0 + \beta_1 X_i)}{1 + \exp(\beta_0 + \beta_1 X_i)}$$

Les observacions X es suposen com a constants conegudes. Alternativament, si les observacions X són aleatòries, $E\{Y_i\}$ és la mitja condicionada als valors de X_i .

2.2.4. Model Múltiple de Regressió Logística:

Com succeeix, amb la regressió lineal, la regressió logística simple pot estendre's a un model més complicat amb més variables explicatives.

De fet, és normal que siguin necessàries més variables explicatives per obtenir una adequada descripció del model i poder fer prediccions útils.

Per passar d'un model simple de regressió logística a un model múltiple, reemplaçem $\beta_0 + \beta_1 x$ de l'equació 1 per $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{p-1} x_{p-1}$.

En aquest punt, simplifiquem les fórmules utilitzant notació vectorial i, definirem els següents dos vectors:

$$\beta^T = [\beta_0 \ \beta_1 \ \beta_2 \ \dots \ \beta_{p-1}] \text{ (dimensió } p \cdot 1)$$

$$X_i^T = [1 \ X_{i1} \ X_{i2} \ \dots \ X_{i,p-1}] \text{ (dimensió } p \cdot 1)$$

Aleshores, amb aquesta notació tenim:

$$\beta^T X_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_{p-1} x_{i,p-1}$$

i, per tant, la funció resposta logística simple (equació 2.2.1) s'estén en una funció resposta logística múltiple:

$$E\{Y\} = \frac{\exp(\beta^T X)}{1 + \exp(\beta^T X)}$$

i, de nou, la forma equivalent és:

$$E\{Y\} = [1 + \exp(-\beta^T X)]^{-1}$$

De manera similar, amb la transformació logit ja coneguda (equació 2.2.2) podem presentar la funció logit resposta com:

$$\pi' = \beta^T X$$

El model de regressió logística múltiple, com abans, pot ser presentat com segueix:

$$E\{Y\} = \pi_i = \frac{\exp(\beta^T X_i)}{1 + \exp(\beta^T X_i)}$$

De nou, les observacions X es consideren constants conegudes. Si les observacions X són aleatòries, $E\{Y\}$ és la mitja condicionada als valors de $X_{i1} X_{i2} \dots X_{i,p-1}$.

Com en el cas de la funció de regressió logística simple, la forma, respecte $\beta^T X$, de la funció de regressió logística múltiple és monòtona i sigmoïdal; cal destacar, que aquesta forma és quasi lineal quan π està compresa entre 0.2 i 0.8.

2.2.5. Estimació dels paràmetres del model de regressió logística:

Els paràmetres β_0, β_1, \dots del model de regressió lineal s'estimen amb el mètode de mínims quadrats que consisteix en trobar els valors estimats b_0, b_1, \dots que fan mínima la suma dels quadrats dels residus de la mostra.

Cal dir, que en el model de regressió lineal el criteri dels mínims quadrats és equivalent al mètode de la de màxima versemblança.

Els paràmetres β_0, β_1, \dots del model de regressió logística s'estimen pel mètode de màxima versemblança ("maximum likelihood") perquè com que els residus són dicotòmics (seguint una distribució de probabilitat binomial) s'obtenen estimacions més eficients que amb el mètode de mínims quadrats.

Així, com que cada observació Y_i és una variable aleatòria de Bernoulli on:

$$\begin{aligned} P(Y_i = 1) &= \pi_i \\ P(Y_i = 0) &= 1 - \pi_i \end{aligned}$$

podem representar la distribució de probabilitat com:

$$f_i(Y_i) = \pi_i^{Y_i} (1 - \pi_i)^{1 - Y_i}; \quad Y_i = 0, 1; \quad i = 1, \dots, n$$

Com que les Y_i observacions són independents, la seva funció de probabilitat conjunta és:

$$g(Y_1, \dots, Y_n) = \prod_{i=1}^n f_i(Y_i) = \prod_{i=1}^n \pi_i^{Y_i} (1 - \pi_i)^{1 - Y_i}$$

En aquest punt, és més fàcil trobar els estimadors de màxima versemblança treballant amb els logaritmes de la funció de probabilitat conjunta:

$$\begin{aligned} \text{Ln } g(Y_1, \dots, Y_n) &= \text{Ln } \prod_{i=1}^n \pi_i^{Y_i} (1 - \pi_i)^{1 - Y_i} = \sum_{i=1}^n \left[Y_i \ln \left(\frac{\pi_i}{1 - \pi_i} \right) \right] + \sum_{i=1}^n \ln(1 - \pi_i) \end{aligned}$$

(equació 2.2.3)

Ja que $E\{Y_i\} = \pi_i$ per una variable binària, si substituïm en $E\{Y_i\} = [1 + \exp(-\beta^T \mathbf{X}_i)]^{-1}$, arribem a:

$$1 - \pi_i = [1 + \exp(\beta^T \mathbf{X}_i)]^{-1}$$

A més, de $\pi' = \ln \left(\frac{\pi}{1 - \pi} \right)$ i $\pi' = \beta^T \mathbf{X}_i$, obtenim: $\ln \left(\frac{\pi}{1 - \pi} \right) = \beta^T \mathbf{X}_i$

I, per tant, l'equació 2.2.3 pot ser expressada com:

$$\ln L(\beta) = \sum_{i=1}^n Y_i (\beta^T X_i) - \sum_{i=1}^n \ln [1 + \exp(\beta^T X_i)]^{-1}$$

Per trobar els valors $\beta_0, \beta_1, \dots, \beta_{p-1}$ que maximitzen $\ln L(\beta)$ s'han d'utilitzar tècniques de búsqueda numèrica.

Aquests valors -estimadors de màxima versemblança- es denoten com b_0, b_1, \dots, b_{p-1} .

Si anomenem \mathbf{b} , al vector dels estimadors de màxima versemblança:

$$\mathbf{b}^T = [b_0 \ b_1 \ \dots \ b_{p-1}]$$

Així, la funció resposta logística prevista i els valors previstos poden ser expressats com:

$$\hat{\pi}_i = \frac{\exp(\beta^T X_i)}{1 + \exp(\beta^T X_i)} = \frac{1}{1 + \exp(-\beta^T X_i)} \quad (\text{equació 2.2.4})$$

2.2.6. Comparació model de regressió lineal – model de regressió logística:

El model de regressió lineal té la forma:

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{p-1} X_{i,p-1} + \varepsilon_i = \mu_i + \varepsilon_i$$

mentre que el model de regressió logística:

$$y_i = \frac{1}{1 + \exp(-\beta^T X)} + \varepsilon_i = \pi_i + \varepsilon_i$$

on ε_i representa l'error residual

Per a completar aquest resum comparatiu:

	Model de regr. lineal	Model de regr. logística
Model sistemàtic \Rightarrow	Funció lineal (recta, pla)	Funció logística (corba)
Residu de (x_i, y_i) \Rightarrow	$\varepsilon_i = y_i - \mu(y/x_i)$	$\varepsilon_i = y_i - \mu(y/x_i) = 1 - \pi_i$
Distribució dels errors \Rightarrow	$E(\varepsilon_i) = 0$ $V(\varepsilon_i) = \sigma^2$ (constant) $\varepsilon_i \approx$ Llei Normal $(0, \sigma^2)$	$E(\varepsilon_i) = 0$ $V(\varepsilon_i) = \pi_x (1 - \pi_x)$ $P(\varepsilon_i) \approx$ Llei Binomial

2.2.7. Deviança del model:

- **Deviança:**

La deviança del model és una mesura de la bondat de l'ajust del model.

La deviança compara la ln-versemblança del model ajustat amb la ln-versemblança del model amb n paràmetres que ajusta les n observacions perfectament (model saturat).

La deviança es comporta com la desviació típica (S) del model de regressió lineal; és a dir, si s'introdueixen en el model més i més variables la deviança millora, tot i que la millora serà mínima si s'afegeixen variables explicatives irrelevantes.

Cal indicar que encara que ens interessarà tenir un model amb la deviance més petita possible, també necessitem que aquest model sigui ho més explicatiu i simple possible; és a dir, no introduïrem variables innecessàriament per reduir la deviance (com hem dit abans, sempre ho faria) i, si tinguéssim dos models amb deviance semblants ens quedariem amb aquell que sigui més fàcil d'explicar.

Abans de posar explícitament la fórmula és necessari fer una mica de treball previ.

Així, es demostrable que la funció de log-versemblança

$$\ln g(Y_1, \dots, Y_n) = \ln \prod_{i=1}^n \pi_i^{Y_i} (1 - \pi_i)^{1 - Y_i} = \sum_{i=1}^n \left[Y_i \ln \left(\frac{\pi_i}{1 - \pi_i} \right) \right] + \sum_{i=1}^n \ln(1 - \pi_i)$$

és maximitza si $\pi_i = Y_i$.

Si a l'estimador de màxima versemblança de π_i pel model saturat el denotem per $\hat{\pi}_{is}$, aleshores $\hat{\pi}_{is} = Y_i$.

Aleshores la deviança pot ser definida com:

$$DEV(X_0, X_1, \dots, X_{p-1}) = 2 \ln L(\hat{\pi}_{1s}, \dots, \hat{\pi}_{ns}) - 2 \ln L(b_0, b_1, \dots, b_{p-1})$$

on:

- $\ln L(b_0, b_1, \dots, b_{p-1})$: és la funció de log-versemblança amb els valors que la maximitzen pel model ajustat.
- $\ln L(\hat{\pi}_{1s}, \dots, \hat{\pi}_{ns})$: és la funció de log-versemblança definida pels estimadors de màxima versemblança de π_i ($\hat{\pi}_{is}$).

Amb això, pel model de regressió logística

$$E\{Y_i\} = \pi_i = \frac{\exp(\beta^T X_i)}{1 + \exp(\beta^T X_i)}$$

la deviance pot ser expressada com:

$$\text{DEV}(X_0, X_1, \dots, X_{p-1}) = -2 \sum_{i=1}^n [Y_i \ln(\hat{\pi}_i) + (1 - Y_i) \ln(1 - \hat{\pi}_i)] \quad (\text{equació 2.2.5})$$

on $\hat{\pi}_i$ és l'i-èssim valor ajustat pel model de regressió logística a l'equació 2.2.4.

És important comentar que, la deviance del model pot ser usada, com hem dit, com un criteri de bondat de l'ajust i, quan més petita és, millor és l'ajust.

▪ Residus de la Deviança:

L'anàlisi de residus per la regressió logística presenta més dificultats que pels models de regressió lineal per que les respostes Y_i només poden prendre els valors 0 i 1.

Per aquest fet, els residus no es distribueixen de forma normal i, per tant, la seva distribució és desconeguda.

Fent gràfics de punts bivariants dels residus enfront els valors ajustats de les variables explicatives no aporten cap informació.

Un tipus usual de residus per la regressió logística son els residus de la deviança.

El residu de la deviança pel cas i , anomenat dev_i , es defineix com l'arrel quadrada de la contribució del i -èssim cas de la suma del model de la deviança (equació 2.2.5):

$$dev_i = \pm \left\{ -2 [Y_i \ln(\hat{\pi}_i) + (1 - Y_i) \ln(1 - \hat{\pi}_i)] \right\}^{1/2}$$

on el signe es positiu quan $Y_i \geq \hat{\pi}_i$ i negatiu quan $Y_i < \hat{\pi}_i$.

2.2.8. Altres usos de la regressió logística:

▪ Transformació Probit:

És un tipus de funció resposta curvilínia que té la mateixa forma que la funció resposta logística mostrada per l'equació 2.2.1.

S'obté per la transformació de la probabilitat π_i mitjançant la funció probit.

$$\text{probit}(\pi_i) = \Phi^{-1}(\pi_i), \text{ on } \Phi \text{ és la funció de distribució de la normal.}$$

El resultat és el model de regressió probit que pot prendre valors dins del rang $-\infty$ a $+\infty$.

La regressió probit és menys flexible que la regressió logística perquè no pot ser estesa per més d'una variable predictora.

▪ Transformació complementària log-log:

Aquesta és altra funció resposta curvilínia amb la mateixa forma que la funció resposta logística.

S'obté per la transformació de la probabilitat π_i mitjançant la funció $\ln[-\ln(1-\pi_i)]$.

Mitjançant aquesta funció transformem la probabilitat π_i (amb un rang de 0 a 1) a un valor que va de $-\infty$ a $+\infty$.

Capítol 3

Base de Dades

3. BASE DE DADES

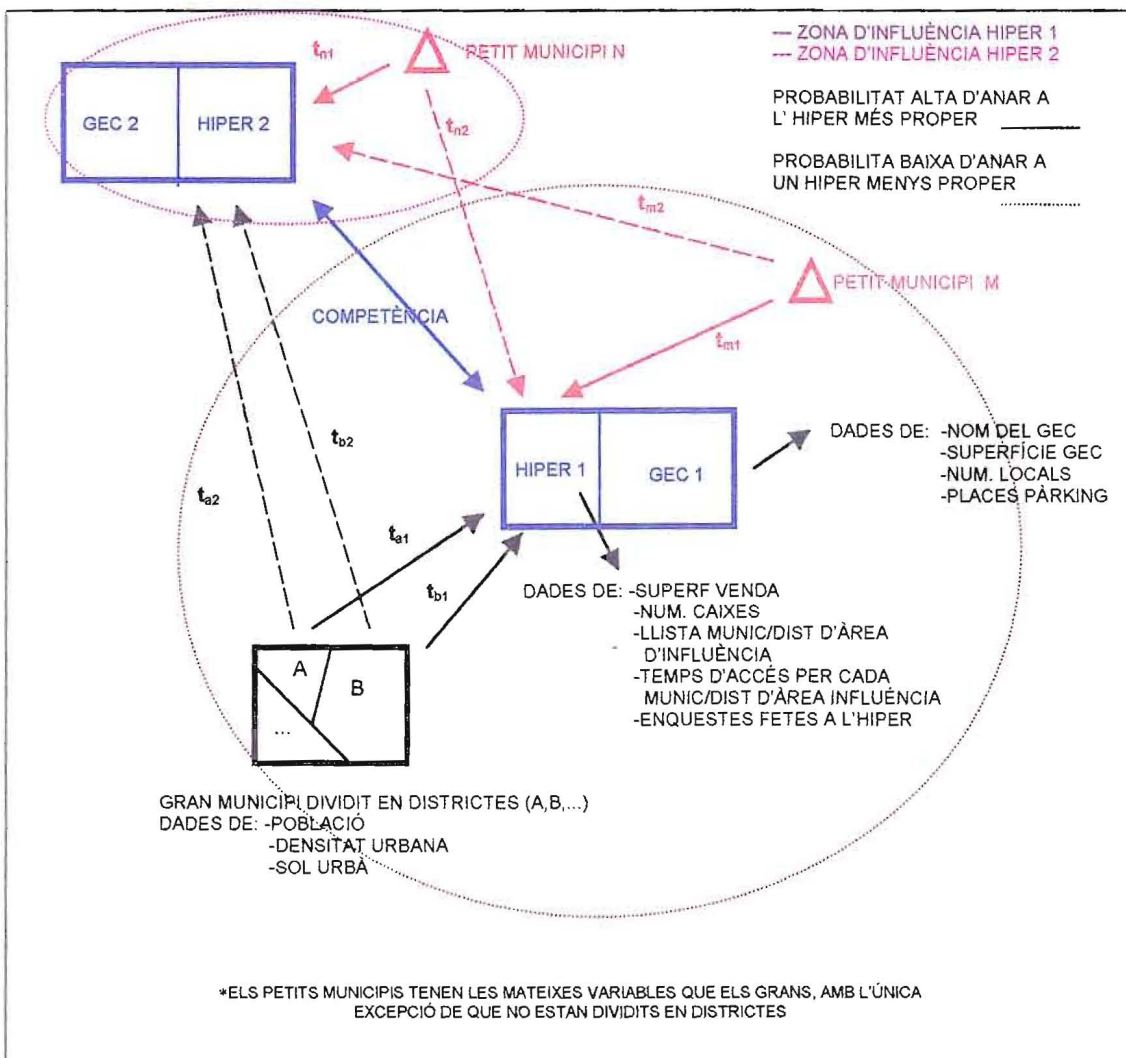
Usarem una base de dades creada a partir d'un recull d'enquestes fetes a les portes dels hipermercats de Grans Equipaments Comercials.

Es van fer 300 enquestes en cadascun dels 45 hipermercats, que es troben en les zones metropolitananes de Madrid, València i Barcelona.

Les variables de les que disposem originalment a la Base de Dades poden ser classificades en cinc grups diferents: dades poblacionals, dades dels GEC's i dels hipermercats, dades de les enquestes, dades de temps i dades de competència.

3.1 DESCRIPCIÓ GRÀFICA DE LA BASE DE DADES

Abans de presentar les variables farem un dibuix esquemàtic per mostrar com són/a què fan referència les variables de que disposem:



(Figura 3.1. Presentació gràfica de com és la situació i algunes de les variables importants)

Sobre l'anterior figura (figura 3.1) cal indicar que s'han exagerat les zones d'influència dels hipermercats. És important comentar que per cada municipi/districte tenim relacions amb hipermercats que estan en un radi aproximat de 20 minuts de distància.

3.2 COMPOSICIÓ DE LA BASE DE DADES

1. Primer grup:

Correspon a les variables que descriuen els municipis/districtes dels que provenen els clients:

- Municipi/districte: nom del municipi/districte
- Població'91: nombre d'habitants del municipi/districte (les dades fan referència al cens de 1991).
- Sòl urbà (Ha): superfície del municipi/districte en hectàrees.
- Denurb: densitat de població del municipi/districte.

Com a exemple d'aquestes dades tenim la següent taula:

Municipi/districte	POBLACIÓ' 91	Sol Urbà (Ha)	Denurb
ABRERA	5464	151,97	3595,4
ALQUÀS	24107	196,00	12299,5
ALBAL	9109	151,01	6032,1
ALCOBENDAS	78825	1345,44	5858,7

(taula 3.1. Exemple de dades de municipis/districtes)

2. Segon grup:

Correspon a les variables que descriuen els hipermercats i els GEC's.

Així pels GEC's tenim:

- Places pk: places de parking que té el Gran Equipament Comercial.
- Num locals: número de locals del que disposa el Gran Equipament Comercial (no es té en compte l'hipermercat).
- SBA m²: superfície bruta (en m²) del Gran Equipament Comercial; aquesta superfície no té en compte la superfície de l'hipermercat, els passadissos, els serveis,...

I pels hipermercats :

- Centre: nom de l'hipermercat.
- Sup. sala ventes m²: superfície (en m²) de la que disposa l'hipermercat.
- Num caixes: número de caixes de l'hipermercat.

Com a exemple d'aquestes dades tenim la següent taula:

Cèntr	supvenm2	num caixes	places pk	núm. locals	SBA m2
Eroski Alcobendas	8800	42	1455	75	30000
Pryca Las Rozas	7300	48	2000	23	8486
Alcampo Alcorcón	10000	68	4290	55	100000

(taula 3.2. Exemple de dades dels hipermercats i dels GEC's)

3. Tercer grup:

Correspon a les variables corresponents a les enquestes fetes a les portes dels hipermercats de cada GEC:

- Importe medio (PTA): indica quina és la despesa en pessetes que destinen els clients procedents d'un determinat municipi/districte en un determinat hipermercat.
- %Procimp: aquesta variable indica quin percentatge de la despesa, sobre la despesa total, és la que representa la que hi dediquen els clients que provenen d'un determinat municipi/districte.
- %Proced: aquesta variable mostra quin és el percentatge de clients que compren en un hipermercat i provenen d'un determinat municipi/districte.

Com a exemple d'aquestes dades tenim la següent taula:

Hipermercat	Municipi/districte	Importe Medio (PTA)	% PROCIMP	% PROCED
Continente Campanar	Extramurs	10553	5,57	4,80
	Campanar	7236	10,18	12,90
	MISLATA	5730	4,64	7,40
	QUART DE POBLET	11600	4,49	3,50
Pryca Gran Turia	L'Eixample	6464	2,93	3,60

(taula 3.3. Exemple de dades de procedents de les enquestes)

4. Quart grup:

La variable que separem a part en aquest grup ens dóna una idea aproximada dels temps d'accés de cada municipi/districte a un GEC.

- Tiempo min: és una estimació del temps d'accés (en minuts) que es triga per anar de cada municipi/districte fins el GEC estudiat. Aquesta és una estimació realitzada tenint en compte les distàncies que els separaven i pel tipus de via per on s'havia de recórrer aquestes distàncies.

Aquesta estimació va ser feta per una mateixa persona per el que s'assumeix homogeneïtat en el sistema de mesura.

Com a exemple:

Hipermercat	Municipi/districte	TIEMPO MIN
Continente Campanar	Extramurs	4,5
	Campanar	3,0
	QUART DE POBLET	14,5
Pryca Gran Turia	L'Eixample	12,0

(taula 3.4. Exemple de dades amb la variable temps d'accés)

5. Cinquè grup:

Aquest grup de variables ens dona una idea aproximada de la competència existent amb altres GEC's; es tenen en compte els 4 GEC's més propers a cada municipi/districte (no s'inclou evidentment el GEC del que se'n vol deduir la despesa).

Per cadascun dels 4 GEC's tenim:

- Hip n: aquest codi indica quin és el primer, segon, tercer o quart hipermercat "més propers" a un determinat municipi/districte (tindrem dades per quatre hipermercats).
- Supv n: superfície de venda (en m²) de l'hipermercat n.
- Hipt n: temps d'accés (en min) entre una població determinada i l'hipermercat n.
- SBA n: superfície bruta (en m²) del Gran Equipament Comercial que conté l'hipermercat n.

Com a exemple d'aquestes variables tenim la següent taula que fan referència a una població i als altres primer i segon hipermercats "més propers":

Municipi/districte	hip1	SBA1	hipt1	hip2	SBA2	hipt2
Eixample	59	53800	7,0	60	31000	7,5
Sants-Montjuïc	61	6500	2,5	60	31000	3,0
Les Corts	60	31000	2,5	61	6500	2,5
Sant Boi de Llobregat	28	32130	6,0	30	43615	8,0
Cornellà de Llobregat	28	32130	2,0	30	43615	8,5

(taula 3.5. Exemple de dades de competència)

6. Noves variables:

Aquest nou grup de variables que definim són les que formen el model definitiu que tenim al Capítol 5.

- Primer: variable categòrica (val 0 o 1). Ha estat creada amb el quocient de variables Població/Temps; per cada hipermercat assignàvem 1 al quocient major i 0 a la resta.

- **% Poblac:** Aquesta variable és un percentatge i per tant està compresa en l'interval [0,100]. Ens indica quin percentatge representa la població d'un municipi/districte respecte del total de població dels municipis/districtes de l'àrea d'influència d'un hipermercat.
- **Comp 1:** variable categòrica (val 0 o 1). Aquesta variable ha estat creada a partir de dues variables de les del grup de competència (quart grup). És una combinació de les variables superfície de venda (supv n) i temps d'accés fins l'hipermercat (Hipt n) que té la següent forma:

$$\sum_{n=1}^4 \text{supv } n / \text{hipt } n$$

Per cada hipermercat, el quocient més gran val 1 i la resta 0.

Aquesta variable de competència intenta reflexar el grau de competència que hi ha entre un o altre hipermercat. Ens dóna un valor que està afavorit pels hipermercats amb més superfície de venda i castiga els estan més lluny dins les rodalies d'un municipi/districte.

Amb la variable intentem tenir en compte el que valoren els clients d'un hipermercat: que prefereixen anar a un hipermercat que està més lluny i que és més gran? o, en canvi, un de més petit i que està més a prop?

- **1/Temps:** Aquesta variable és l'invers de la variable del temps (tiempo min).

Capítol 4

Anàlisi exploratòria de les dades

4. ANÀLISI EXPLORATÒRIA DE LES DADES:

4.1 ANÀLISI UNIVARIANT:

En aquest apartat fem una descripció gràfica de les variables de la base de dades.

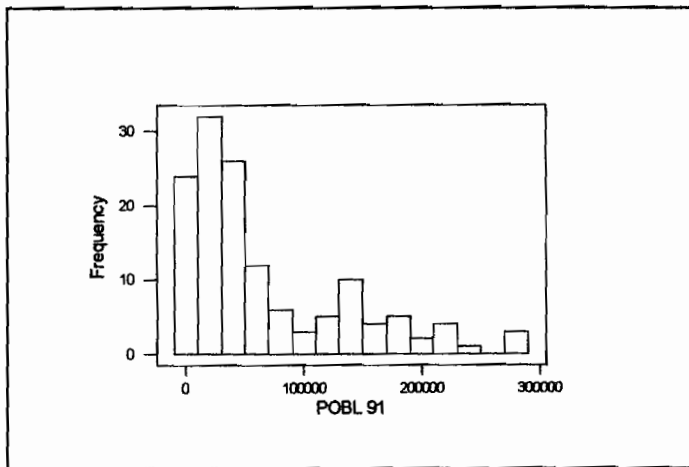
A partir del anàlisi univariant ens podem fer una idea de com és distribueixen les variables de la base de dades.

A més a més, depurarem qualsevol dada incorrecta que creiem que no hi ha d'estar.

4.1.1. Variables de Població:

Per cada població tenim el nom, la superfície en hectàrees i la seva població l'any 1991, segons el cens. Hi ha una tercera variable que és la densitat de població (número d'habitants per quilòmetre quadrat). Aquesta última variable es defineix com una combinació de població i de superfície.

Les variables d'aquest grup es caracteritzen per la asimetria de la distribució deguda a la diferència entre municipis / districtes.



(Figura 4.1.1. Histograma de la variable població)

És molt important mencionar, que com s'observa en l'anterior figura, cap municipi/districte supera els 300.000 habitants, quan es coneix que, per exemple, Barcelona, Madrid i València superen de llarg aquesta xifra.

Aquest fet s'explica perquè sovint hi ha GEC's dins les grans ciutats i per evitar grans diferències en els càlculs les tres capitals han estat fraccionades en districtes. D'aquesta manera tenim en la base de dades les ciutats i pobles de les zones metropolitanes de Madrid, Barcelona i València i, també, els districtes de les tres grans capitals.

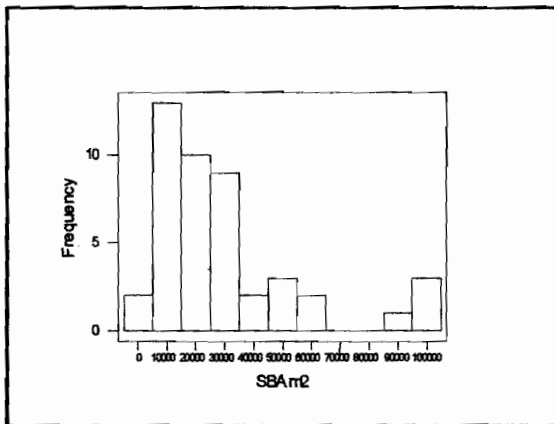
4.1.2. Variables dels Grans Equipaments.

En aquest apartat es tracten diverses variables que poden servir per caracteritzar una gran superfície comercial: places de pàrking, número de locals, superfície bruta en m², etcètera.

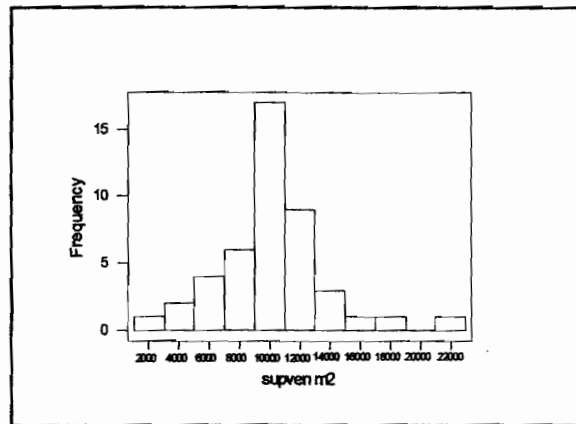
També n'hi ha dues que caracteritzen el tamany de l'hipermercat: número de caixes i superfície de la sala de ventes en m².

Cal esmentar que a la base de dades original, dels Hipercor no es disposa del número de locals que tenien i, a més, un d'ells no deia el número de caixes de que disposa el seu hipermercat.

A continuació presentem l'anàlisi gràfica de les dues variables que creiem més destacades, la superfície bruta en lloguer dels GEC's (SBA m²) i el de la superfície de venda dels hipermercats (supven m²), respectivament.



(Figura 4.1.2. Histograma variable superfície dels GEC's)



(Figura 4.1.3. Histograma variable superfície dels hipers)

En el cas de la superfície de venda dels hipermercats (figura 4.1.3) s'observa una clara simetria d'aquesta superfície entorn dels aproximadament 11.000 m² de venda.

En canvi, pel que fa a la superfície bruta dels GEC's (figura 4.1.2) es veu que les dades són clarament asimètriques i que la majoria de GEC's no passen dels 35.000 m².

Això, en conjunt, sembla indicar que hi ha dos tipus de GEC's diferenciats: els que es centren en l'alimentació i que tenen un hipermercat molt gran i poques botigues, i uns altres centrats més en l'oferta no alimentària (per tant hipermercats menors) i que disposen de moltes botigues.

4.2 ANÀLISI BIVARIANT

Com hem vist en l'anàlisi de la base de dades, les variables podem separar-se en diferents grups depenent de si són variables que fan referència als municipis/districtes, als hipermercats i GEC's, al comportament dels clients i a la competència.

Per tant, en aquest punt volem analitzar les possibles relacions bivariants entre les variables explicatives dels tres primers grups. Posteriorment, també analitzarem les relacions entre les variables explicatives del model (que triem) i la variable resposta.

Aquesta anàlisi ens pot aportar més informació per comprendre millor el problema.

4.2.1. Dades de municipis/districtes:

Les dades de població de les que disposem fan referència a municipis/districtes de les àrees metropolitanas de Madrid, Barcelona i València.

De cada municipi/districte tenim la població (dades del cens del 1991), la superfície urbana (en hectàrees) i una combinació d'aquestes anteriors que és la densitat urbana (quantitat de població per km²).

Com es comprova en la matriu de gràfics bivariants de l'annex, cap de les relacions existents entre les variables que fan referència a municipis/districtes (sol urbà vs població, densitat urbana vs. població i densitat urbana vs. sol urbà) ens permet assolir cap conclusió destacada; per aquest fet, hem decidit deixar-los en l'annex.

4.2.2. Dades dels hipermercats i dels Grans Equipaments Comercials:

En aquest punt el nostre objectiu és intentar trobar alguna relació destacada entre les variables que fan referència als hipermercats i els GEC's.

Com que no és el mateix un hipermercat que el Gran Equipament Comercial que el conté, farem una anàlisi per separat de les variables que corresponen a un o altre.

A la base de dades les variables referides a hipermercats són la superfície de l'hipermercat (supvenm2) i el número de caixes.

Per un altre costat, les variables referides als GEC's són el número de places de pàrking, el número de locals del GEC (sense tenir en compte l'hipermercat) i la seva superfície bruta (SBAm2).

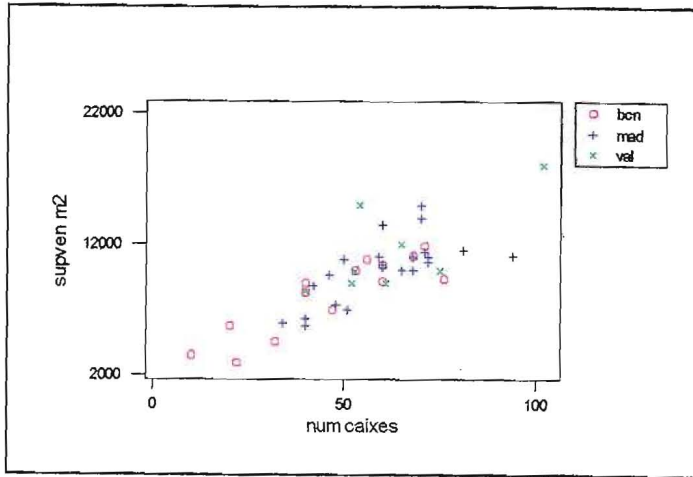
Com podem comprobar en l'annex, les correlacions més destacades amb aquestes variables són les següents:

- supvenm2 vs. num caixes (pels hipermercats).
- SBAm2 vs. num locales (pels GEC's).
- SBAm2 vs. places pk (pels GEC's).

Les anteriors correlacions són força intuïtives, sembla que sigui normal que quan augmenta la superfície de l'hipermercat augmenti el número de caixes; també, és força acceptable que els GEC's majors són els que tenen més locals i més places de pàrking.

A continuació presentem breument cadascuna de les tres anteriors relacions.

▪ **Supven m2 vs. num caixes:**



(figura 4.2.1. Número de caixes vs. Superfície de venda segons zona metropolitana)

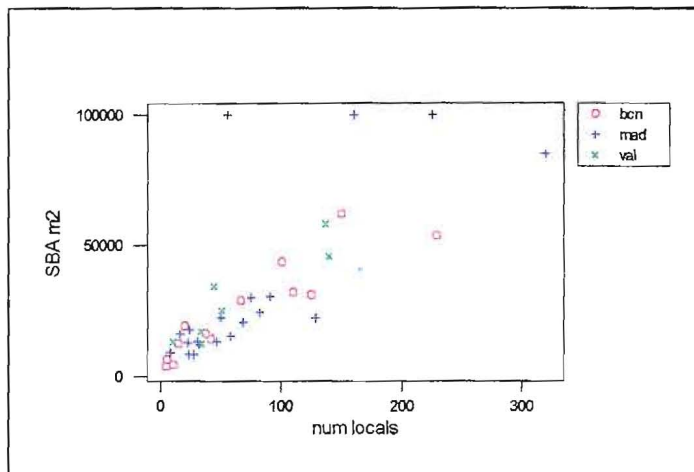
Mitjançant aquesta relació ens podem fer una certa idea de la grandària de l'hipermercat i la relació que existeix amb el número de caixes.

En la figura 4.2.1 es comprova la correlació existent entre aquestes dues variables.

És una relació lògica -directament proporcional- que ens mostra que quan augmenta la superfície de l'hipermercat augmenta el número de caixes.

Podem acceptar que els hipermercats més grans han de tenir més caixes.

▪ **SBA m2 vs. num locals:**



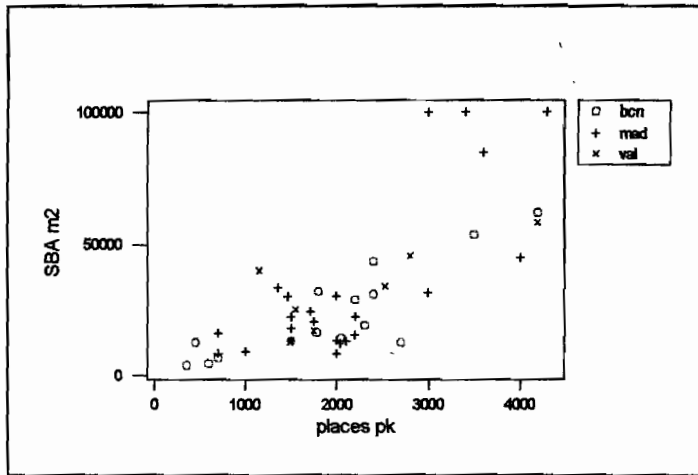
(figura 4.2.2. Número de locals vs. superfície bruta segons zona metropolitana)

Amb aquesta relació que fa referència als GEC's també podem tenir una idea aproximada del seu tamany i de la seva capacitat per albergar més o menys locals.

Es comprova que els GEC's amb més superfície són els que més locals contenen.

Cal indicar, però, que no apareixen les dades de cap Hipercor (doncs no es tenien les xifres dels locals).

▪ **SBA m2 vs. places pk:**



(figura 4.2.3. Places de pàrking vs. superfície bruta segons zona geogràfica)

En aquesta figura, com en les anteriors, volem mostrar una manera d'expressar la grandària d'un GEC; si en l'anterior es feia a partir del número de locals, aquí ho fem a partir del número de places de pàrking.

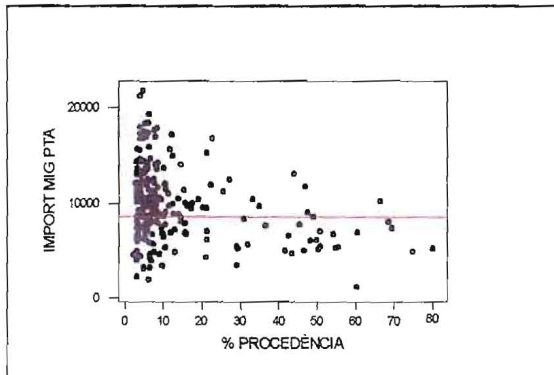
A la figura 4.2.3 s'observa la correlació que hi ha entre la superfície dels GEC's i el número de places de pàrking dels mateixos; de fet, com s'observa en l'annex, aquesta és la relació més feble de les tres presentades.

4.2.3. Dades del comportament del consum:

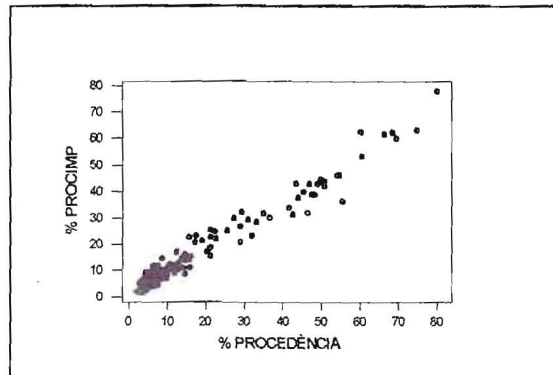
A la base de dades tenim variables que, bàsicament, fan referència al consum i a la procedència dels clients que compren en els hipermercats dels GEC's.

Per tant, en aquest punt el que volem és intentar veure quina és la relació entre la procedència dels clients i la seva despesa en els hipermercats.

Les dues següents figures són les més destacades en aquest punt de l'anàlisi.



(figura 4.2.4. Relació despesa mitja-procedència)



(figura 4.2.5. Relació %Procimp-%Procedència)

La figura 4.2.4 compara la despesa mitja dels clients amb la procedència.

Aquesta figura és molt important perquè observem que per municipis/districtes amb un percentatge de procedència baix la variació de despesa és molt alta, mentre que per municipis/districtes amb un percentatge de procedència alt la despesa és pràcticament la mateixa.

Així podriem dir que el valor esperat de la despesa (**línia vermella**) és aproximadament constant; i, per tant, que, sigui quina sigui la procedència dels clients, la despesa mitja és la mateixa.

La figura 4.2.5 mostra, com es comprova a l'annex, una de les correlacions més altes entre dues variables existents en la base de dades.

Aquesta relació és una de les més importants en el present projecte.

Sovint, preguntar a la gent el que es gasta és un fet una mica incòmode i moltes vegades la gent no contestarà o si ho fa, la seva resposta pot ser molt subjectiva; a més, pot passar com en el cas dels Hipercor que no és deixi fer aquesta pregunta als enquestadors.

L'alta correlació entre ambdues variables és normal doncs %procedència fa referència al percentatge d'un municipi/districte que compra en un hipermercat determinat respecte el total de població que hi compra, mentre que %procimp és similar a l'anterior i també contempla la despesa en aquell hipermercat.

De fet, al mancar dades i perquè és una apreciació subjectiva no podem modelar sobre la despesa de la població en un hipermercat; per raons semblants tampoc podem fer-ho amb %procimp, però si que podem i, de fet farem, intentar modelar a partir del % procedència.

Capítol 5

Metodologia

5. METODOLOGIA

Com hem comentat anteriorment, l'objectiu del projecte passa per modelar la proporció de clients que assistiran a un hipermercat en funció del municipi/districte al que pertanyen.

Si es modela la proporció de clients en certa manera s'està modelant la facturació procedent d'un municipi/districte a un hipermercat en concret; com s'havia vist en l'anàlisi bivariànt.

Com s'ha comentat en el capítol 4 (apartat 4.2.3), la proporció de facturació i la proporció d'assistència tenen una correlació molt forta i positiva.

Això és important perquè és menys violent preguntar de quin municipi procedeix un client d'un hipermercat que no quants diners s'ha gastat. A més a més, preguntar directament a un client l'import de la seva compra pot provocar una resposta subjectiva subjecta a biaix.

Un altre entrebanc, ja comentat també, és que hi ha hipermercats que no permeten preguntar al client els diners que s'ha gastat.

Com es veu en l'anàlisi bivariànt l'import mig té una esperança constant. Per tant només caldrà conèixer la proporció de clients que provenen d'un municipi determinat per conèixer, mitjançant l'import mig, la estimació de la proporció de facturació d'un municipi en concret.

Per tant resumint, el model que hem de trobar ens ha de estimar una proporció.

Recordem que es van fer, aproximadament, 300 enquestes a cada hipermercat.

Es té per cada municipi/districte, el número d'enquestes realitzades, llavors és evident que la suma de totes les enquestes realitzades ha de ser aproximadament 300; però es dona el cas que hi ha clients que provenen d'altres municipis/ districtes que no són de la zona d'influència del GEC, aquests casos es van treure i no es van tenir en compte.

Per tant a partir del número d'enquestes realitzades es pot treure fàcilment la proporció de clients que provenen d'un municipi en concret. Un cop es té la proporció de clients, és fàcil estimar la proporció de facturació que es deu a un municipi o districte.

El que realment importarà saber es quantes enquestes de les realitzades, són d'un dels municipis o districtes de la zona d'influència

$$N_{i,j} = \pi_{i,j} * 300$$
$$\forall i,j$$

Aquesta formula ens explica la manera d'obtenir la proporció de clients en funció del número d'enquestes contestades per cada municipi/ districte; on:

- N fa referència al número d'enquestes,
- el subíndex i fa referència al municipi/districte i
- el subíndex j fa referència a l'hipermercat.

Per tant $N_{i,j}$ fa referència al número d'enquestes contestades per clients del municipi i -èssim a l'hipermercat j -èssim.

El símbol $\pi_{i,j}$ fa referència a la proporció de clients del municipi i -èssim que compren en l'hipermercat j -èssim.

Així doncs, i per comoditat per treballar amb el paquet estadístic Minitab, s'ha decidit modelar el número d'enquestes realitzades a cada municipi/districte i no pas la proporció de clients que assisteixen a cada GEC.

En aquest cas particular ho modelarem mitjançant l'acció *Binary Logistic Regression*, del paquet estadístic Minitab. No cal dir que per la naturalesa de la variable no es pot modelar mitjançant un model de regressió lineal. Tot això ja ha estat explicat en el capítol de teoria on s'explica la diferència entre la regressió lineal i la regressió logística.

Capítol 6

Procés de Modelat

6. PROCÉS DE MODELAT

6.1 COM HA DE SER EL MODEL?

Com s'ha comentat en altres apartats el que es vol trobar és un model que ens estimi de manera satisfactòria quina és la proporció de gent que compra en l'hipermercat d'un GEC. L'estimació es farà a partir de les enquestes realitzades en els hipermercats.

Prioritzarem alguns criteris per escollir el model.

Un primer criteri serà trobar un model que, evidentment ajusti bé tots els casos que tenim; caldria veure si el mateix model pot ser extrapolable a altres zones metropolitanas.

El segon dels criteris que hem tingut en compte és que el model sigui fàcil d'interpretar, que sigui intuïtiu i de fàcil comprensió.

Buscarem que el model no tingui moltes variables i minimitzarem, sempre que sigui possible, el número de transformacions de variables.

Resumint, volem aconseguir un model que expliqui de manera satisfactòria la proporció de procedència de clients d'un municipi en un hipermercat i que a part sigui intuïtiu i senzill de comprendre.

6.2 PRIMERA PART DE LA MODELITZACIÓ

Si separem la modelització en tres parts. La primera seria una part molt llarga on s'intenta modelar moltes variables alhora i on la dificultat, a la hora d'interpretar, és elevada. En aquesta primera fase, es pot considerar que 'jugàvem' amb les dades i estudiàvem les relacions més importants entre elles. En aquesta fase es van trobar models que no complien el primer requisit i molt menys el segon dels requisits.

En aquesta part del modelatge ja es comença a entreveure les variables que seran decisives en el model i les variables de les que es podran prescindir. Per exemple, en aquest punt del modelatge es va veure que la variable sol urbà no la necessitaríem per res ja que no aportava informació alguna.

Una altre exemple de variable que veiem que no entra en el model seria la del número de pàrkings de l'hipermercat. Aquesta variable, en contra de l'exemple anterior, ens va sorprendre ja que nosaltres creiem que era un factor rellevant alhora de decidir a on anar a comprar. Aquesta variable, com es comprova en l'Annex A2, està força correlacionada amb la variable Superfície Bruta del GEC (SBA) i, per tant, si tingués que entrar alguna de les dues, seria la variable SBA i no pas el número de places de pàrking.

Durant aquesta part del modelatge es va veient quines variables són les realment importants i les que no són. Es va fent una selecció de les variables que seran determinants en el model.

6.3 SEGONA PART DE LA MODELITZACIÓ

La segona part de la modelització es caracteritza per la necessitat de trobar noves variables que ens ajudin a escollir el model segons el segon criteri. En aquest apartat es va intentar formar moltes noves variables amb diferent grau de dificultat.

En una primera fase es va intentar crear variables indicadores de les diferents comunitats autònomes. Per tant, tenint en compte que hi havia tres comunitats autònomes es van crear dos noves variables indicadores.

Aquestes variables es van multiplicar per les diferents variables original; volíem veure com actuaven aquestes en les respectives zones metropolitanes.

En el següent exemple es pot observar que les variables població i temps actuen de manera diferent a València que a Madrid.

És lògic que amb transformacions com aquestes el model expliqui més, però, al mateix temps, el model perd senzillesa.

Cal recordar que aquest és un exemple significatiu d'aquesta fase però que es van provar altres models.

```
Results for: BASEDADESMM.MTW
Binary Logistic Regression: %PROCENQ; N ENQ versus DENSITAT_1; SUPVEN; ...
Link Function: Logit
Logistic Regression Table
```

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Constant	-2,12065	0,03776	-56,16	0,000			
DENSITAT	7,9119E-06	1,0508E-06	7,53	0,000	1,00	1,00	1,00
SUPVEN	-2,794E-05	3,3724E-06	-8,29	0,000	1,00	1,00	1,00
pobval	-8,081E-06	1,0982E-06	-7,36	0,000	1,00	1,00	1,00
pobmad	-3,349E-07	1,7870E-07	-1,87	0,061	1,00	1,00	1,00
ltempva	0,8274	0,1902	4,35	0,000	2,29	1,58	3,32
ltempma	4,9502	0,1033	47,91	0,000	141,20	115,32	172,90

(Taula 6.1. Sortida de Minitab per regressió logística amb noves variables)

A part d'aquesta transformació van intentar altres transformacions i nous càlculs, que en ajudaven a interpretar el model. Dintre d'aquestes transformacions podem citar a mode d'exemple la variable 'població partit per temps' o 'densitat partit per temps', etcètera. Moltes d'aquestes variables no van afectar als resultats.

Durant aquesta segona fase, vam intentar diversos molts models i noves transformacions.

A continuació, presentem les quatre variables que intervenen d'alguna forma en el model final. Totes elles són variables que provenen de transformacions d'unes altres variables de la base de dades original.

- **Variable Competència:**

Vam veure que necessitàvem una variable que ens indiqués el grau de competència que tenia l'hipermercat dintre de la seva zona d'influència. Aprofitant que coneixíem els quatre hipermercats més propers a un municipi, vam definir cinc indicadors de competència.

Aquests indicadors ens estimaran, en certa manera, el grau de competència existent dins d'una zona delimitada pels 5 hipermercats més propers (l'estrictament més proper al municipi i els altres quatre més propers).

Tots els indicadors castigaven el fet de tenir un hipermercat de gran superfície a la zona. Es pot entendre que com més gran sigui un hipermercat és més probable que absorbeixi clients d'altres hipermercats. Sembla lògic creure que si hi ha hipermercats petits i un hipermercat gran a la zona, serà més probable que la gent vagi a aquest hipermercat gran i no als petits.

Un altre factor per mesurar la competència, vam decidir que fos, el temps mínim d'accés de cada municipi/districte a l'hipermercat al que es referia. Per tant s'entén que la gent assisteix als hipermercats més propers.

Aquests cinc indicadors de competència eren semblants entre ells però diferents alhora. Per exemple un utilitzava els temps al quadrat en canvi un altre simplement els temps. Un cop analitzat i discutit el bo i el dolent de cada indicador ens vam quedar amb el següent indicador:

$$\text{comp}_{i,j} = \sum_{k=1}^{k=4} \frac{\text{superf_vent}_k}{\text{temps}_{i,k}}$$

S'ha definit el grau de competència com el sumatori pels quatre hipermercats més propers, per cada hipermercat, de la fracció entre superfície de venda i el temps d'accés des de el municipi fins a l'hipermercat.

▪ **Variable Primer:**

Una altra nova variable, que es va calcular va ser la variable primer. Aquesta variable és categòrica i ens indica quin és el municipi/ districte més important dintre de la zona d'influència de l'hipermercat.

Per tant prendrà el valor 1 si és el municipi més important i 0 en cas contrari. Només queda explicar com s'ha calculat la importància de cada municipi dintre de la seva zona.

Es va decidir que els municipis més grans serien més importants que els municipis petits, però també havia de ser un factor determinant el temps d'accés. Per tant el municipi més gran i més proper a l'hipermercat, en principi, hauria de ser el municipi amb major número d'enquestes.

El que es va fer, va ser calcular el quocient entre població i temps mínim per cada un dels municipis de la zona. El municipi amb un quocient més gran se li va assignar un 1 i als altres un 0.

▪ **Variable %Població:**

Finalment es va crear una variable que anomenarem, % Població. Aquesta variable ens indica el percentatge de població d'un municipi/districte en concret en funció de la població total de la zona d'influència. Ho il·lustrarem amb l'exemple de l'hipermercat del GEC (Continent de Campanar):

S'observa que hi ha 9 municipis/ districtes a la seva zona d'influència. Cada municipi/districte té una població determinada i per tant es pot calcular la població del conjunt de la zona (377481). Per cada un dels municipis es calcula la proporció de població que aporta, i és aquesta proporció la que ens interessarà per trobar bons models.

Nom del municipi/districte	Població l'any 91	% Població
Extramurs	52713	14%
Campanar	31178	8%
La Saldia	48824	13%
L'Olivereta	51351	14%
Patraix	46296	12%
Rascanya	44117	12%
Benicalap	36941	10%
Mislata	38808	10%
Quart de Poblet	27253	7%
Població total	377481	100%

(Taula 6.2.Exemple de taula per la variable % Població)

▪ Variable 1/Temps:

Aquesta va ser la variable més fàcil de crear.

Simplement vam prendre la variable Temps (que venia originalment en la base de dades) i vam fer la seva inversa (dividir 1 per la variable Temps) per crear la nova variable.

Al fer la seva inversa el que volem és "castigar" els hipermercats que són més lluny d'un municipi/districte i "primar" els que estan més prop d'un municipi/districte.

Prenem aquesta variable perquè el temps d'accés d'un municipi/districte a un GEC creïem que és un factor que influeix alhora d'anar o no l'hipermercat del GEC en qüestió.

Aquestes són les transformacions més importants que s'han fet. Encara que s'hagin fet altres transformacions o càlculs, només hem indicat les que creïem més importants i que tindrem en compte en el model.

6.4 TERCERA PART DE LA MODELITZACIÓ

En aquesta tercera fase el que s'ha fet és una selecció entre els models que creiem millors, aplicant els dos criteris exposats al principi del capítol.

En definitiva, al final del procés, ens vam quedar amb tres models per escollir. A continuació exposarem aquests tres models:

Model 1		Model 2		Model 3	
Deviança	4988	Deviança	5781	Deviança	4953
Concordant Percent	72,70%	Concordant Percent	71,00%	Concordant Percent	73,00%
Variables	Coefficients	Variables	Coefficients	Variables	Coefficients
Constant	-0,66647	Constant	-3,45591	Constant	-1,50305
Primer	0,82058	Primer	1,07928	Primer	0,79034
Competència	-8,52E-05	Competència	-6,87E-05	Competència	-8,97E-06
% Població	0,0324647	Ln(% Població)	0,4788	Ln(% Població)	0,56311
Ln (1/T ²)	0,45433	1/T ²	5,6878	Ln (1/T)	0,92996

(Taula 6.3. Taula comparativa dels tres models finals)

Tenim un tercer criteri per escollir un model que consisteix en triar el que tingui la deviança més petita i un percentatge de concordància major.

Així doncs, resumint per triar un model teníem en compte, primer, els següents criteris:

1. Que el model que sigui el més homogeni possible per tota la base de dades.
2. Que el model sigui intuïtiu i de fàcil interpretació.
3. Si tenim diversos models molt similars, triar el que tingui menor deviança i major percentatge de concordància.

Un punt important és que, a part d'aquests tres criteris, hem de mirar els gràfics que mostren la distribució dels residus per comprovar que no mostrin cap patró i que estiguin distribuïts aleatòriament.

Tenint en compte tot això ens vam quedar amb el tercer model.

El primer criteri no variava molt entre els tres models i es podia considerar igual en els tres. La manera de veure-ho va ser amb el gràfic dels residus (veure punt A.5 de l'annex).

Pel que fa al segon criteri, cal esmentar les següents igualtats:

$$\ln(1/T) = -\ln T$$

$$\ln(1/T^2) = -2 \ln T$$

Capítol 7

Resultats i interpretació

7. RESULTATS I INTERPRETACIÓ

Al final del capítol anterior havíem arribat a un model satisfactori, recordem que era un model amb quatre variables. Aquestes quatre variables havien estat calculades a partir de les variables inicials:

- **Competència:** Mesura la competència de l'hipermercat en funció de la distància i superfície dels quatre hipermercats més propers.
- **Primer:** Mesura el municipi/ districte més important dintre de la zona d'influència de l'hipermercat.
- **Ln (% Població):** Mesura el percentatge de població que representa el municipi/ districte dintre de la zona d'influència.
- **Ln (1/T):** Mesura la distància d'un municipi/ districte a l'hipermercat. Passa a dividir ja que volem 'castigar' els municipis/districtes més allunyats. Sembla clar que com més lluny estigui el municipi/ districte menys enquestes s'obtindran d'ell.

Nota: En les dues últimes variables s'utilitza el logaritme neperià per motius de millorar la distribució dels residus i reduir la curvatura d'aquests. (Mirar punt A.5 de l'annex).

Model definitiu.	
Deviança	4953
Concordant Percent	73,00%
Variables	Coefficients
Constant	-1,50305
Primer	0,79034
Competència	-8,97E-05
Ln (1/T)	0,92996
Ln (% Població)	0,56311

(Taula 7.1. Taula resum dels coeficients del model definitiu)

En la taula anterior podem veure la estimació dels coeficients, per tant, es pot començar a interpretar alguna cosa.

La primera variable del model ens diu que el fet de ser el municipi/districte **més** important de la zona d'influència de l'hipermercat, **implica** que hi hagi **més** proporció de clients. En el cas que no sigui el primer, es pot comprovar que no varia.

La segona de les variables té una estimació del coeficient negatiu. El que implica que a **més** competència **menys** proporció de clients. Recordem que estem parlant en termes mitjos i no vol dir que incrementar la competència d'un hipermercat disminueixi el número de clients sinó que pot ser que augmentin, però en terme mig disminueix. S'observa que el valor del coeficient és petit però això no vol dir que l'efecte sigui petit, ja que la variable competència pot prendre valors molt grans.

Semblava lògic que hipermercats amb molta competència tinguessin menys proporció de clients (l'inversa també es certa).

Al cap i a la fi, tenint en compte el capítol de Metodologia, més enquestes implica més percentatge de clients i per tant més percentatge de ingressos provinents d'un municipi/ districte determinat a un hipermercat en concret.

La tercera variable del model és el logaritme neperià de $1/T$. S'observa un valor positiu del coeficient, ens indica que a més temps (distància) menys proporció de clients número d'enquestes i a l'invers.

La quarta variable que entraria en el model final és el logaritme neperià de la proporció de població d'un municipi/districte dintre de la zona d'influència d'un hipermercat. Aquesta variable com té un coeficient positiu ens indica que com més gent hi hagi al municipi/districte (com més gran) més clients provindran d'aquell municipi/ districte.

A continuació mostrem la sortida del paquet estadístic Minitab.

Binary Logistic Regression: PROCENQ; N ENQ versus primero; Competència; ...

Link Function: Logit

Response Information

Variable	Value	Count
PROCENQ	Success	10412
	Failure	71188
N ENQ	Total	81600

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Constant	-1,50305	0,06294	-23,88	0,000			
primero	0,79034	0,03051	25,91	0,000	2,20	2,08	2,34
Competèn	-8,969E-05	6,1614E-06	-14,56	0,000	1,00	1,00	1,00
ln pobla	0,56311	0,01768	31,84	0,000	1,76	1,70	1,82
ln 1/T	0,92996	0,02270	40,97	0,000	2,53	2,42	2,65

Log-Likelihood = -27242,581

Test that all slopes are zero: G = 7823,788; DF = 4; P-Value = 0,000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	5274,751	267	0,000
Deviance	4953,613	267	0,000
Hosmer-Lemeshow	306,834	8	0,000

Table of Observed and Expected Frequencies:

(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value Group	Success		Failure		Total
	Observed	Expected	Observed	Expected	
1	400	228,3	8000	8171,7	8400
2	455	420,1	7945	7979,9	8400
3	513	517,1	7887	7882,9	8400
4	519	608,2	7881	7791,8	8400
5	574	694,5	7826	7705,5	8400
6	967	787,1	7433	7612,9	8400
7	841	957,1	7559	7442,9	8400
8	1021	1274,1	7379	7125,9	8400
9	2295	2221,5	6105	6178,5	8400
10	2827	2704,0	3173	3296,0	6000
				-----	81600

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	541181511	73,0%	Somers' D	0,47
Discordant	191690511	25,9%	Goodman-Kruskal Gamma	0,48
Ties	8337434	1,1%	Kendall's Tau-a	0,10
Total	741209456	100,0%		

En la sortida que dona el Minitab, s'observa per cada variable el coeficient, el valor de la z, el p_valor de significació, etcètera.

Després arribem a la part on es defineix la bondat del model. En aquest apartat ens fixem sobretot en la Deviance, es més baixa que en altres models. Però no només ens hem de fixar en aquest indicador ja que com més variables s'afegeix, més baixa. Encara que si ens servirà per comparar models amb el mateix número de variables.

Els altres dos indicadors es comporten de manera similar però no ens fixem en ells.

Al final del model hi ha una taula que també ens interessa, aquest apartat ens indica el percentatge de concordància. Ens interessarà que aquest sigui el major possible. Fixar-nos que ens indica el percentatge d'explicació. Per tant si el model si el valor fos 0%, voldria dir que tenim una relació de soroll blanc, en canvi si el valor fos del 100% tindriem un model determinista. En el nostre cas no està malament, tenim un 73% explicat.

En aquest punt del capítol caldria veure els residus del model.

En el gràfic dels residus (veure Annex A.5.3) s'observa que no són dolents, però **tampoc** són perfectes. El principal defecte d'aquest gràfic és que els residus semblen **presentar** una certa tendència.

Creiem que a pesar d'aquest fet, i amb les dades que teníem, el model que presentem **és** el millor que es pot fer.

- Veiem que el model final té la següent formula matemàtica:

$$E\{\% \text{Procedència}\} = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4)} =$$

$$= \frac{\exp(-1.5 + 0.7 \cdot \text{primer} - 8.96 \cdot 10^{-5} \cdot \text{Competència} + 0.93 \cdot \text{Ln}(T^{-1}) + 0.56 \cdot \text{Ln}(\% \text{Població})}{1 + \exp(-1.5 + 0.7 \cdot \text{primer} - 8.96 \cdot 10^{-5} \cdot \text{Competència} + 0.93 \cdot \text{Ln}(T^{-1}) + 0.56 \cdot \text{Ln}(\% \text{Població})}$$

Capítol 8

Conclusions

8. CONCLUSIONS

8.1 CONCLUSIONS DEL MODEL

S'ha trobat un model que explica de manera satisfactòria la proporció de clients que provindran d'un municipi/ districte determinat a l'hipermercat d'interès. Aquest model utilitza com a *inputs* la població del municipi/districte, el temps mínim d'accés a l'hipermercat i un indicador de la competència entre els hipermercats de la zona.

En contra del que es pensava inicialment, per estimar la procedència, no fem referència a les característiques estructurals dels GEC's (num. Places pàrking, num locals,...).

Un cop es té l'estimació de la proporció de clients, serà relativament fàcil conèixer la proporció de facturació que assumeix el municipi/ districte en l'hipermercat.

Hem de recordar que a l'anàlisi exploratòria es va veure que hi havia una correlació significativa -que val 0.983- entre la proporció de facturació per municipi/districte i la proporció de clients. (Veure punt A.3 dels Annexos).

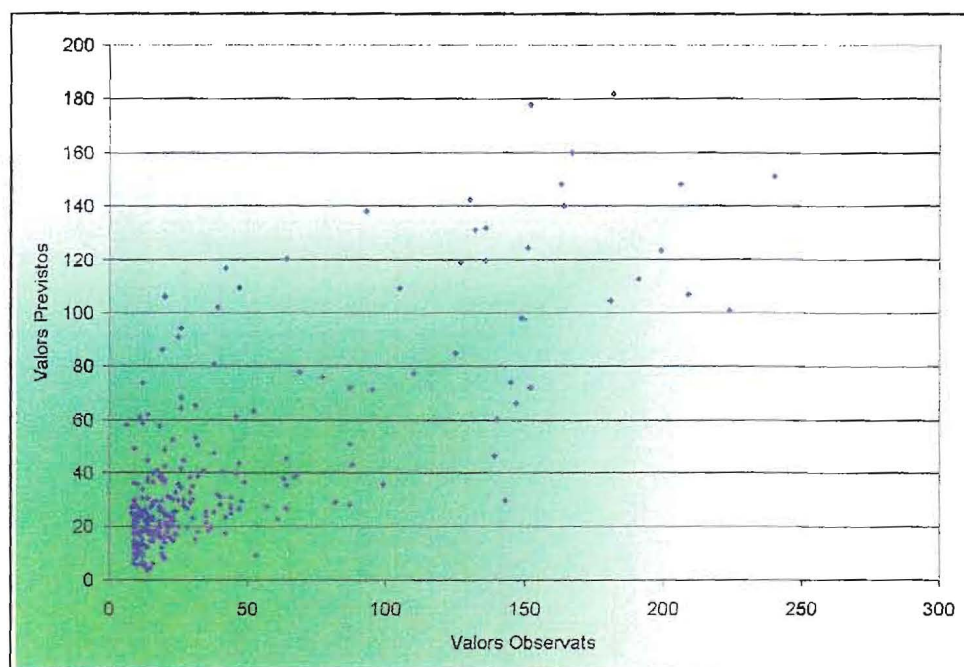
Només ens falta recordar, que com s'ha dit en l'apartat anterior, el model tendeix a subestimar els municipis/ districtes que assumeixen moltes enquestes i a sobrestimar els municipis/ districtes que tenen poques enquestes.

Això és així per la forma que té la fórmula dels Residus de la Deviança (veure punt 2.2.7 de la Introducció Teòrica). Quan el valor real és major que el valor previst pel model estem subestimant, mentre que quan és el valor previst major que el valor real estem sobreestimant.

Tot i arribar a un model satisfactori, sabem que podríem tenir-ne un de millor introduïnt variables no recollides en la base de dades. Seria interessant haver preguntat en les enquestes variables biodemogràfiques, socio-econòmiques, ... com, per exemple, edat o grups d'edat, tamany de nucli familiar i professió.

Nota: En l'apartat A.6 de l'annex s'ha diferenciat el model per cada àrea metropolitana (Barcelona, Madrid i València).

Amb els valors previstos pel model i els valors observats podem crear la següent gràfica:



(Figura 8.1. Gràfic bivariant de Valors Observats vs. Valors Previstos)

Com s'observa en la figura 8.1 i en les taules del punt A.7 dels Annexos, s'observa una gran variabilitat entre els valors observats i els valors previstos pel model.

Per valors observats molt grans el model dona valors previstos petits. Això succeeix perquè tot i que el model és el millor per les dades que tenim hi ha molta variabilitat entre municipis.

Possiblement, darrera dels hàbits de compra dels clients dels municipis/districtes que tenim hi han variables que no estan registrades –les desconeixem- i són importants per determinar millor la proporció de gent que compra en un hipermercat.

8.2 EXEMPLE.

En aquest apartat es mostrarà un petita taula per un hipermercat en concret (Continente Campanar) i els municipis/ districtes de la seva zona d'influència.

Aquesta taula també conté les variables explicatives del model, la variable resposta -amb els seus valors observats i esperats-.

HIPERMERCAT	MUNIC / DISTR	Primer	LN 1/T	LN %població	Competència	PROCENQ	V. PREVISTOS
CONTIN CAMPANAR	Extramurs	1	-1,50408	2,63651	8022,9	14	62
	Campanar	0	-1,09861	2,11136	8813,3	39	32
	La Saidia	0	-1,79176	2,55987	5063,1	12	30
	L'Olivereta	0	-1,60944	2,61033	13403,3	36	18
	Patraix	0	-2,07944	2,50671	6283,1	21	21
	Rascanya	0	-2,0149	2,4585	4474,4	12	25
	Benicalap	0	-1,79176	2,28097	5092	11	26
	MISLATA	0	-2,19722	2,33028	7329,7	22	16
	QUART DE POBLET	0	-2,67415	1,97681	4486,1	11	11

(Taula 8.2 Exemple en l'hipermercat Continente Campanar)

Cal recordar que PROCENQ \equiv Valors Observats

Bibliografia

BIBLIOGRAFIA

Per la realització del present projecte ens hem bassat principalment en:

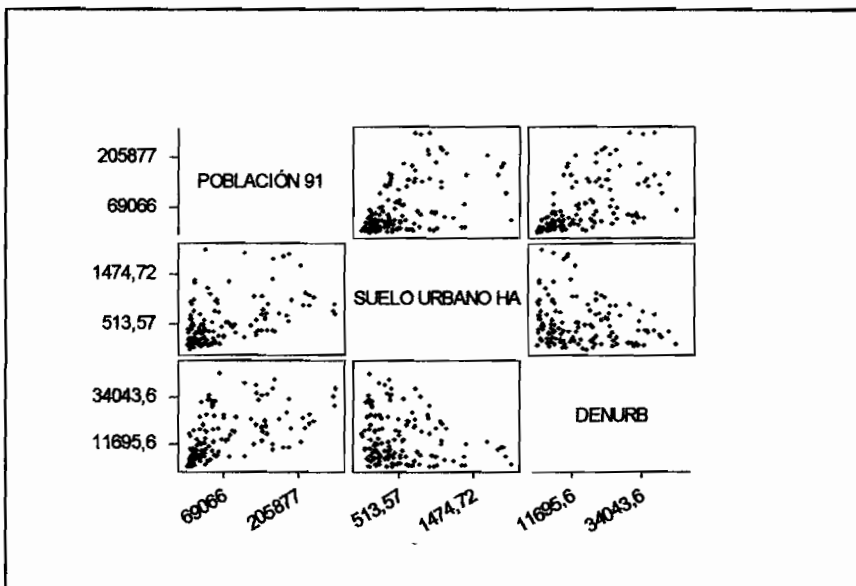
- Doménech, José M.; Sarriá, Alfonso. (1997). *Unidad Didáctica 13. Modelo de regresión logística*.
Barcelona: Signo. Col.: *Análisis Multivariante en Ciencias de la Salud. Modelos de Regresión*, 51 p.
- John Neter, Michael H. Kutner, Christopher J. Nachstein, William Wasserman. (1996). Logistic Regression, Poisson Regression and Generalized Linear Models.
En: *Applied Linear Statistical Models*. Chicago: Irwin, p. 567 – 627.
- Josep Ginebra (2000). *Apunts de Models Lineals*.
- Universitat Politècnica de Catalunya (1997). *Aprendre a aprendre: Com presentar un treball acadèmic*. Barcelona: la Universitat. Full solt.
- Universitat Politècnica de Catalunya (1997). *Aprendre a aprendre: Referències bibliogràfiques per a documents impresos*. Barcelona: la Universitat. Full solt.

Annexos

ANNEXOS

Tot i que presentem els annexos com una sola unitat, s'ha de tenir clar que tenim bàsicament 4 parts: la primera la formen els 4 primer apartats que corresponen a les correlacions i el gràfics bivariants de les diferents variables; la segona part la formen la comparativa de diversos models i el model final; la tercera part la forma la presentació del model final separat per zona metropolitana; i, finalment, la quarta part la forma la taula construïda a partir de la base de dades presentant les diferents variables del model i els valors observats i els valors previstos.

A.1 MATRIU DE GRÀFICS BIVARIANTS I CORRELACIONS PER VARIABLES DE POBLACIÓ:

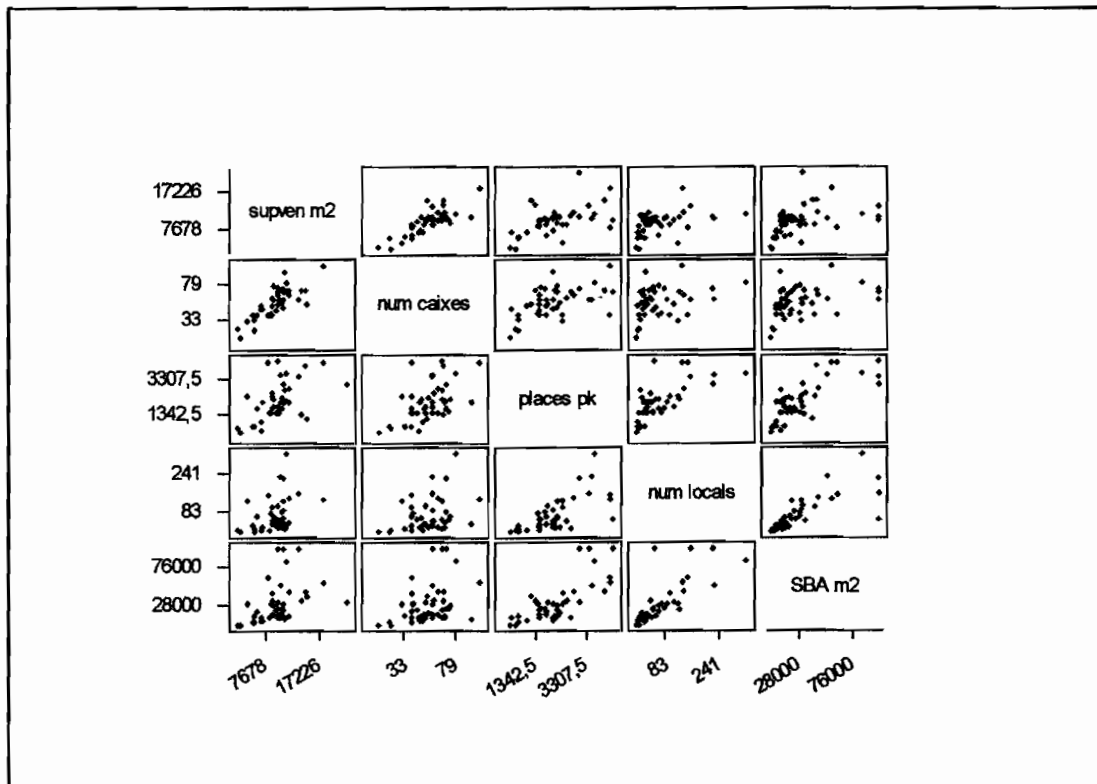


(figura A.1. Matriu de gràfics bivariants de variables poblacionals)

- Per aquest grup de variables també presentem les correlacions existents entre elles:

	POBLACIÓ	SOL URBÀ
SOL URBÀ	0,492 0,000	
DENS URB	0,565 0,000	-0,230 0,007

Cell Contents: Pearson correlation
P-Value

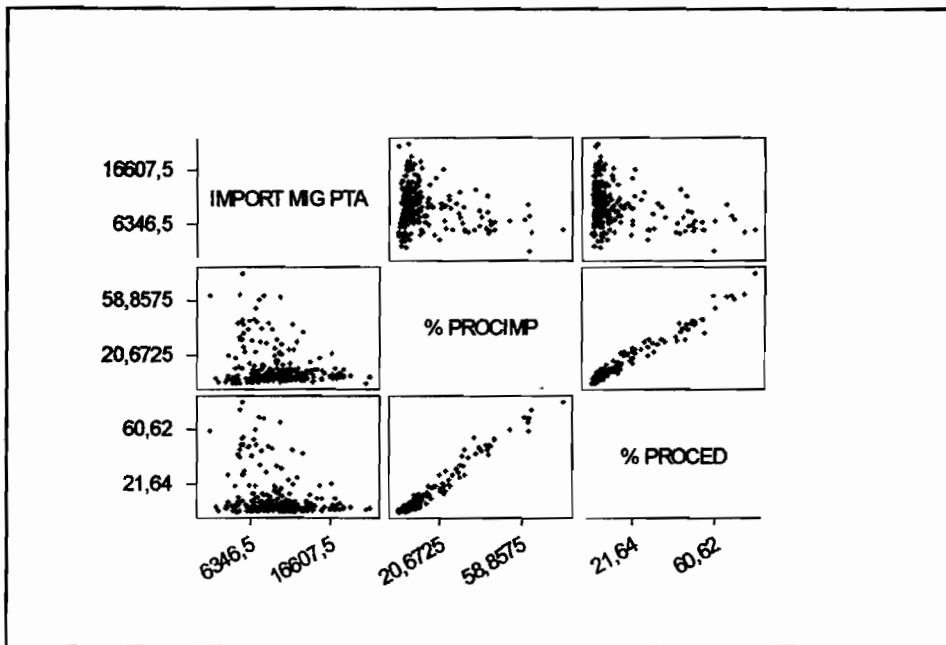
A.2 MATRIU DE GRÀFICS BIVARIANTS I CORRELACIONS PER VARIABLES**DELS HIPERMERCATS I GEC'S:**

(figura A.2. Matriu de gràfics bivariants de var's d'hipermercats i de GEC's)

- Per aquest grup de variables també presentem les correlacions existents entre elles:

	supven m2	num caixes	places pk	num locals
num caixes	0,799 0,000			
places pk	0,548 0,000	0,543 0,000		
num locals	0,400 0,011	0,344 0,030	0,675 0,000	
SBA m2	0,385 0,009	0,365 0,015	0,753 0,000	0,772 0,000

Cell Contents: Pearson correlation
P-Value

A.3 MATRIU DE GRÀFICS BIVARIANTS I CORRELACIONS PER VARIABLES DE CONSUM:

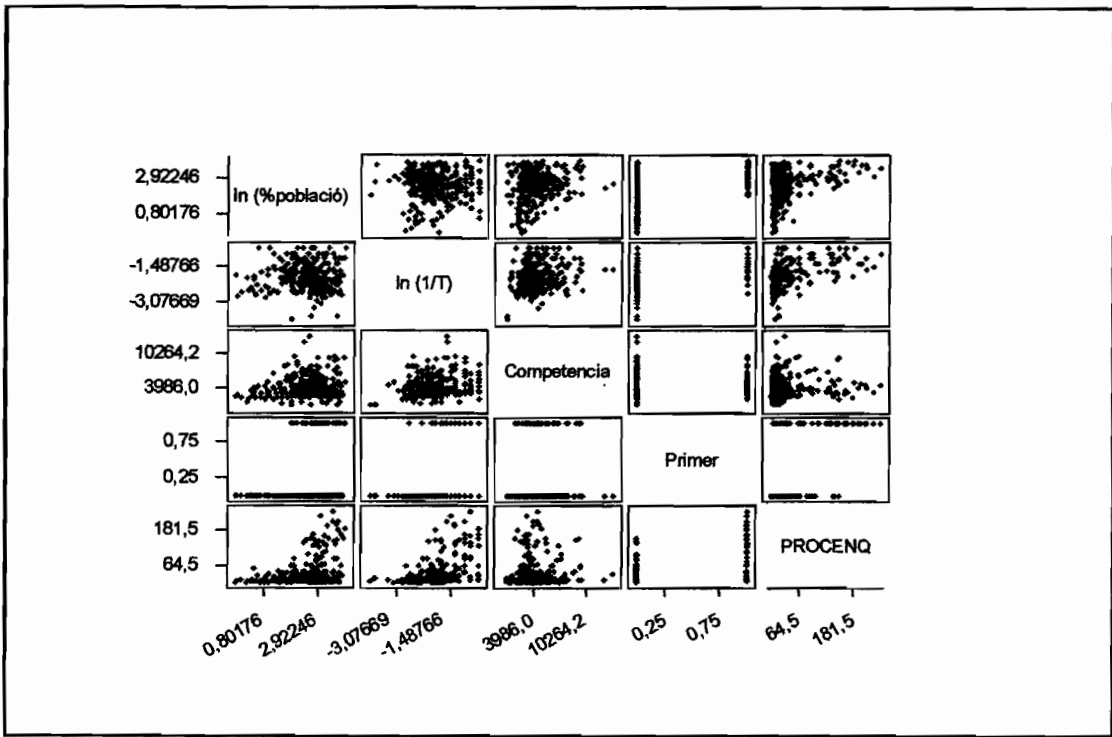
(figura A.3. Matriu de gràfics bivariants de variables de consum)

- Per aquest grup de variables també presentem les correlacions existents entre elles:

	% PROCED	% PROCIM
% PROCIM	0,983 0,000	
IMPORT M	-0,273 0,000	-0,215 0,001

Cell Contents: Pearson correlation
P-Value

A.4 MATRIU DE GRÀFICS BIVARIANTS I CORRELACIONS PER VARIABLES DEL MODEL:



(figura A.4. Matriu de gràfics bivariants de variables de consum)

- Per aquest grup de variables també presentem les correlacions existents entre elles:

	Primer	ln (1/T)	ln (%pob	Competen
ln (1/T)	0,459	0,000		
ln (%pob	0,317	-0,015	0,800	
Competen	0,106	0,241	0,177	0,003
PROCENQ	0,638	0,503	0,347	0,020
	0,000	0,000	0,000	0,740

Cell Contents: Pearson correlation
P-Value

A.5 MODELS FINALS I SORTIDES DE MINITAB:

En aquest apartat de l'annex incloem les sortides de Minitab dels 3 models que comparàvem en l'apartat 6.4 del Capítol 6.

Per cadascun dels tres models també incloem el gràfic bivariant que mostra la relació entre els valors de la variable resposta i els residus de la deviança.

A.5.1. Primer Model.

Variables: Primer, Competencia, %Població, ln(1/T2)

Results for: BASEDAESM.MTW

Binary Logistic Regression: PROCENQ; N ENQ versus Primer; Competencia; ...

Link Function: Logit

Response Information

Variable	Value	Count
PROCENQ	Success	10412
	Failure	71188
N ENQ	Total	81600

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Constant	-0,66647	0,05455	-12,22	0,000			
Primer	0,82058	0,03014	27,22	0,000	2,27	2,14	2,41
Competen	-8,519E-05	6,1573E-06	-13,84	0,000	1,00	1,00	1,00
% Poblac	0,0324647	0,0009730	33,37	0,000	1,03	1,03	1,03
Ln (1/T2	0,45433	0,01115	40,73	0,000	1,58	1,54	1,61

Log-Likelihood = -27259,781

Test that all slopes are zero: G = 7789,389; DF = 4; P-Value = 0,000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	5342,606	267	0,000
Deviance	4988,011	267	0,000
Hosmer-Lemeshow	177,301	8	0,000

Table of Observed and Expected Frequencies:

(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value Group	Success		Failure		Total
	Observed	Expected	Observed	Expected	
1	365	310,8	8035	8089,2	8400
2	475	438,0	7925	7962,0	8400
3	505	514,1	7895	7885,9	8400
4	597	585,7	7803	7814,3	8400
5	652	665,3	7748	7734,7	8400
6	766	746,5	7634	7653,5	8400
7	993	938,7	7407	7461,3	8400
8	999	1288,2	7401	7111,8	8400
9	1995	2168,6	6405	6231,4	8400
10	3065	2756,2	2935	3243,8	6000

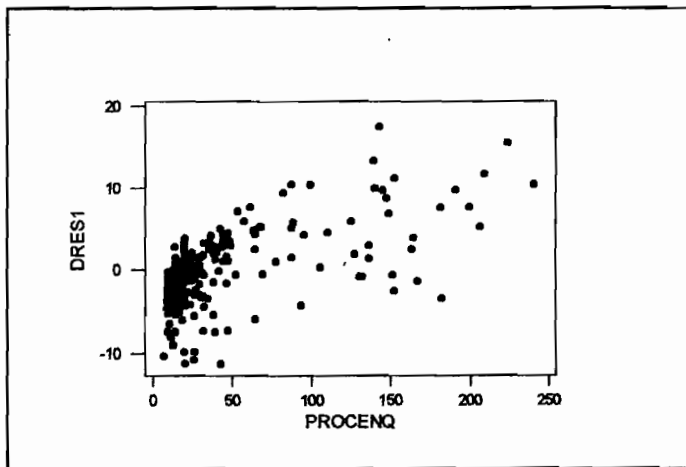
81600

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	539138894	72,7%	Somers' D	0,47
Discordant	193573094	26,1%	Goodman-Kruskal Gamma	0,47
Ties	8497468	1,1%	Kendall's Tau-a	0,10
Total	741209456	100,0%		

Per aquest primer model el gràfic bivariant que compara els valors observats per la variable resposta (PROCENQ) i els residus de la deviança és:



(Figura A.5.1. Gràfic bivariant de Valors Observats per la resposta vs. Residus deviança)

A.5.2. Segon Model.

Variables: Primer, Competencia, Ln(%Població), 1/T2

Binary Logistic Regression: PROCENQ; N ENQ versus Primer; Competencia; ...

Link Function: Logit

Response Information

Variable	Value	Count
PROCENQ	Success	10412
	Failure	71188
N ENQ	Total	81600

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Constant	-3,45591	0,05293	-65,29	0,000			
Primer	1,07928	0,02933	36,80	0,000	2,94	2,78	3,12
Competen	-6,865E-05	6,1298E-06	-11,20	0,000	1,00	1,00	1,00
Ln (% Po	0,47880	0,01695	28,25	0,000	1,61	1,56	1,67
1/TEMPS2	5,6878	0,1862	30,55	0,000	295,25	204,98	425,27

Log-Likelihood = -27656,715

Test that all slopes are zero: G = 6995,521; DF = 4; P-Value = 0,000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	6317,201	267	0,000
Deviance	5781,880	267	0,000
Hosmer-Lemeshow	48,078	8	0,000

Table of Observed and Expected Frequencies:

(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

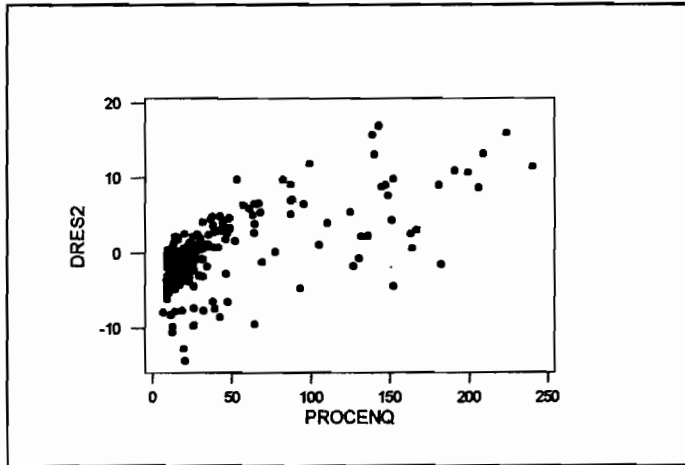
Value Group	Success		Failure		Total
	Observed	Expected	Observed	Expected	
1	429	330,8	7971	8069,2	8400
2	516	490,1	7884	7909,9	8400
3	538	593,1	7862	7806,9	8400
4	613	661,6	7787	7738,4	8400
5	706	715,9	7694	7684,1	8400
6	798	786,6	7602	7613,4	8400
7	821	886,3	7579	7513,7	8400
8	1094	1104,4	7306	7295,6	8400
9	2193	2174,3	6207	6225,7	8400
10	2704	2668,9	3296	3331,1	6000
				-----	81600

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	526574420	71,0%	Somers' D	0,44
Discordant	204021020	27,5%	Goodman-Kruskal Gamma	0,44
Ties	10614016	1,4%	Kendall's Tau-a	0,10
Total	741209456	100,0%		

Pel segon model el gràfic bivariant que compara els valors observats per la variable resposta (PROCENQ) i els residus de la deviança és:



(Figura A.5.2. Gràfic bivariant de Valors Observats per la resposta vs. Residus deviança)

A.5.3. Model Definitiu.

Variables: Primer, Competencia, Ln(%Població), ln(1/T)

Binary Logistic Regression: PROCENQ; N ENQ versus Primer; Competencia; ...

Link Function: Logit

Response Information

Variable	Value	Count
PROCENQ	Success	10412
	Failure	71188
N ENQ	Total	81600

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Constant	-1,50305	0,06294	-23,88	0,000			
Primer	0,79034	0,03051	25,91	0,000	2,20	2,08	2,34
Competen	-8,969E-05	6,1614E-06	-14,56	0,000	1,00	1,00	1,00
Ln (% Po	0,56311	0,01768	31,84	0,000	1,76	1,70	1,82
Ln (1/T)	0,92996	0,02270	40,97	0,000	2,53	2,42	2,65

Log-Likelihood = -27242,581

Test that all slopes are zero: G = 7823,788; DF = 4; P-Value = 0,000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	5274,751	267	0,000
Deviance	4953,613	267	0,000
Hosmer-Lemeshow	306,834	8	0,000

Table of Observed and Expected Frequencies:

(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value Group	Success		Failure		Total
	Observed	Expected	Observed	Expected	
1	400	228,3	8000	8171,7	8400
2	455	420,1	7945	7979,9	8400
3	513	517,1	7887	7882,9	8400
4	519	608,2	7881	7791,8	8400
5	574	694,5	7826	7705,5	8400
6	967	787,1	7433	7612,9	8400
7	841	957,1	7559	7442,9	8400
8	1021	1274,1	7379	7125,9	8400
9	2295	2221,5	6105	6178,5	8400
10	2827	2704,0	3173	3296,0	6000

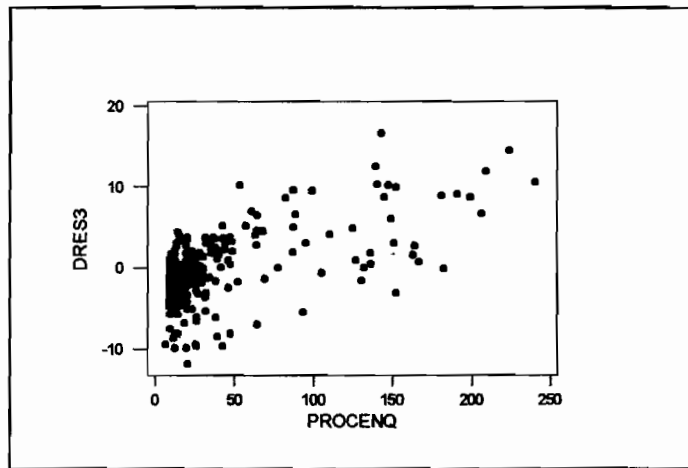
					81600

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	541181511	73,0%	Somers' D	0,47
Discordant	191690511	25,9%	Goodman-Kruskal Gamma	0,48
Ties	8337434	1,1%	Kendall's Tau-a	0,10
Total	741209456	100,0%		

Pel model definitiu, dels tres candidats finals, el gràfic bivariant que compara els valors observats per la variable resposta (PROCENQ) i els residus de la deviança és:



(Figura A.5.3. Gràfic bivariant de Valors Observats per la resposta vs. Residus deviança)

A.6 SORTIDES DE MINITAB DEL MODEL DEFINITIU PER ZONA METROPOLITANA

Com a complement de l'estudi realitzat en el present projecte hem dividit la Base de Dades en les tres zones metropolitanes que teníem: València, Madrid i Barcelona.

Hem aplicat en cadascuna d'elles el model final de regressió logística anteriorment trobat; recordem que aquest model té per variables explicatives: Primer, Competència, Ln (%Població) i Ln (1/T).

Per resumir la informació que tindrem en les sortides de Minitab presentem la següent taula:

	València	Madrid	Barcelona
Coefficients			
Constant	-2.0574	-0.3299	-1.70334
Primer	0.00477	1.05775	0.47477
Competencia	$-9.06 \cdot 10^{-5}$	$-9.13 \cdot 10^{-5}$	$-5.27 \cdot 10^{-5}$
Ln (1/T)	0.81255	1.06338	1.1080
Ln (% Poblacio)	0.63628	0.29114	0.68332
Deviança	410.617	1952.783	1758.412
% Concordant	60.3 %	76.0 %	71.7 %

(Taula A.6.1. Coeficients del model i Indicadors de dispersió)

A.6.1. Sortida de Minitab per la Zona Metropolitana de València:

Link Function: Logit

Response Information

Variable	Value	Count
%PROCENQ	Success	1546
	Failure	20054
N ENQ	Total	21600

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Constant	-2,0574	0,1667	-12,35	0,000			
Primer	0,00477	0,08852	0,05	0,957	1,00	0,84	1,20
Competen	-9,063E-05	0,0000138	-6,54	0,000	1,00	1,00	1,00
Ln (1/T)	0,81255	0,06163	13,19	0,000	2,25	2,00	2,54
Ln (%pob	0,63628	0,05849	10,88	0,000	1,89	1,68	2,12

Log-Likelihood = -5384,821

Test that all slopes are zero: G = 362,639; DF = 4; P-Value = 0,000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	428,798	67	0,000
Deviance	410,617	67	0,000
Hosmer-Lemeshow	41,597	7	0,000

Table of Observed and Expected Frequencies:

(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value Group	Success		Failure		Total
	Observed	Expected	Observed	Expected	
1	86	63,9	2314	2336,1	2400
2	113	101,9	2287	2298,1	2400
3	140	122,4	2260	2277,6	2400
4	169	146,1	2231	2253,9	2400
5	125	162,0	2275	2238,0	2400
6	158	178,7	2242	2221,3	2400
7	195	202,4	2205	2197,6	2400
8	179	223,4	2221	2176,6	2400
9	381	345,1	2019	2054,9	2400

					21600

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	18708427	60,3%	Somers' D 0,24
Discordant	11205127	36,1%	Goodman-Kruskal Gamma 0,25
Ties	1089930	3,5%	Kendall's Tau-a 0,03
Total	31003484	100,0%	

A.6.2. Sortida de Minitab per la Zona Metropolitana de Madrid:

Link Function: Logit

Response Information

Variable	Value	Count
PROCENQ	Success	5449
	Failure	30251
N ENQ	Total	35700

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Constant	-0,3299	0,1066	-3,10	0,002			
Primer	1,05775	0,04373	24,19	0,000	2,88	2,64	3,14
Competen	-9,130E-05	9,0686E-06	-10,07	0,000	1,00	1,00	1,00
Ln (1/T)	1,06338	0,03444	30,88	0,000	2,90	2,71	3,10
Ln (%pob	0,29114	0,02361	12,33	0,000	1,34	1,28	1,40

Log-Likelihood = -12722,312

Test that all slopes are zero: G = 5060,978; DF = 4; P-Value = 0,000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	2004,379	114	0,000
Deviance	1952,783	114	0,000
Hosmer-Lemeshow	211,191	8	0,000

Table of Observed and Expected Frequencies:

(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value Group	Success		Failure		Total
	Observed	Expected	Observed	Expected	
1	168	129,9	3432	3470,1	3600
2	172	187,8	3428	3412,2	3600
3	227	234,8	3373	3365,2	3600
4	245	266,9	3355	3333,1	3600
5	343	299,6	3257	3300,4	3600
6	358	351,3	3242	3248,7	3600
7	301	410,7	3299	3189,3	3600
8	566	610,3	3034	2989,7	3600
9	1510	1212,6	2090	2387,4	3600
10	1559	1745,0	1741	1555,0	3300
					35700

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	125260017	76,0%	Somers' D	0,53
Discordant	37079217	22,5%	Goodman-Kruskal Gamma	0,54
Ties	2498465	1,5%	Kendall's Tau-a	0,14
Total	164837699	100,0%		

A.6.3. Sortida de Minitab per la Zona Metropolitana de Barcelona:

Link Function: Logit

Response Information

Variable	Value	Count
PROCENQ	Success	3417
	Failure	20883
N ENQ	Total	24300

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Constant	-1,70334	0,09858	-17,28	0,000			
Primer	0,47477	0,05363	8,85	0,000	1,61	1,45	1,79
Competen	-5,278E-05	0,00001466	-3,60	0,000	1,00	1,00	1,00
Ln (1/T)	1,10080	0,04492	24,50	0,000	3,01	2,75	3,28
Ln (%pob	0,68332	0,03074	22,23	0,000	1,98	1,86	2,10

Log-Likelihood = -8719,548

Test that all slopes are zero: G = 2296,512; DF = 4; P-Value = 0,000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	1911,299	76	0,000
Deviance	1758,412	76	0,000
Hosmer-Lemeshow	221,855	7	0,000

Table of Observed and Expected Frequencies:
(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value Group	Success		Failure		Total
	Observed	Expected	Observed	Expected	
1	128	81,9	2572	2618,1	2700
2	159	143,9	2541	2556,1	2700
3	161	181,6	2539	2518,4	2700
4	173	215,2	2527	2484,8	2700
5	396	275,9	2304	2424,1	2700
6	354	354,0	2346	2346,0	2700
7	310	447,0	2390	2253,0	2700
8	478	611,8	2222	2088,2	2700
9	1258	1105,6	1442	1594,4	2700
					----- 24300

Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	51179856	71,7%	Somers' D	0,45
Discordant	18989856	26,6%	Goodman-Kruskal Gamma	0,46
Ties	1187499	1,7%	Kendall's Tau-a	0,11
Total	71357211	100,0%		

A.7 TAULA DE LA BASE DE DADES

HIPERMERCAT	MUNIC / DISTR	Primer	LN 1/T	LN %població	Competencia	PROCENQ	V. PREVISTOS
CONTINENTE CAMPANAR	Extramurs	1	-1,50408	2,83651	8022,9	14	62
	Campanar	0	-1,09861	2,11136	8813,3	39	32
	La Saldia	0	-1,79176	2,55987	5063,1	12	30
	L'Olivereta	0	-1,60944	2,61033	13403,3	36	18
	Patraix	0	-2,07944	2,50671	6283,1	21	21
	Rascanya	0	-2,0149	2,4585	4474,4	12	25
	Benicalap	0	-1,79176	2,28097	5092	11	26
	MISLATA	0	-2,19722	2,33028	7329,7	22	16
	QUART DE POBLET	0	-2,67415	1,97681	4486,1	11	11
	PRYCA GRAN TURIA	L'Eixample	0	-2,48491	2,26383	6764,4	11
Extramurs		0	-1,79176	2,381	7218,4	10	23
L'Olivereta		1	-0,69315	2,35482	6379,3	20	106
Patraix		0	-1,94591	2,25119	4946,7	30	23
Jesús		0	-2,48491	2,26849	4263,4	20	15
Quatre Carreres		0	-2,63906	2,63959	7837,7	14	12
TORRENT		0	-2,25129	2,43432	2891,6	23	23
ALQUÀS		0	-2,0149	1,59864	3313,6	10	18
XIRIVELLA		0	-0,91629	1,67819	5237,5	16	40
MISLATA		0	-1,38629	2,07476	3927,5	20	37
CONTINENTE ALFAFAR	ALDAIA	0	-2,07944	1,52296	2899,1	11	17
	PICANYA	0	-1,94591	0,46885	3998,6	12	10
	Quatre Carreres	0	-2,60269	3,3269	7836,2	32	18
	Pobles del Sud	0	-1,50408	1,98908	5873,3	18	27
	TORRENT	0	-2,70805	3,12163	4053,4	9	20
	PAIORTA	0	-2,07944	1,85244	5021,7	15	16
CONTINENTE ALFAFAR	CATARROJA	0	-2,30259	2,1061	3697,9	14	17
	ALBAL	0	-2,63906	1,31271	3132,1	10	9

HIPERMERCAT	MUNIC / DISTR	Primer	LN 1/T	LN %població	Competencia	PROCENQ	V. PREVISTOS
	ALFARAR	1	-0,69315	2,09857	8160,3	19	86
	BENETÚSSER	0	-1,38629	1,73404	8160,3	18	22
	SEDAVÍ	0	-0,91629	1,19048	6641,3	16	28
	SILLA	0	-2,56495	1,90832	4750,1	10	11
PRYCA L'ELIANA	BÉTERA	0	-2,0149	2,26675	2627,9	28	26
	RIBA-ROJA	0	-2,19722	2,2686	2467,2	22	23
	PATERNA	1	-2,74084	3,74345	4972,7	32	50
	POBLA DE VALBONA	0	-1,09861	2,08062	3514,5	38	48
	L'ELIANA	0	-1,79176	2,07653	2557,3	48	29
	LLÍRIA	0	-2,14007	2,62423	2449,4	16	29
	BENEGUASIL	0	-2,19722	2,15898	2387	23	22
ALCAMPO ALBORAYA	La Saldia	0	-2,35138	2,90177	5852,6	10	21
	Pla del Real	0	-1,79176	2,46496	4717,7	23	30
	Poblets Marítims	0	-2,3979	3,08502	2474,5	9	29
	Algirós	1	-2,0149	2,72279	4279,6	18	58
	Benimaclet	0	-2,19722	2,34185	4460,4	21	20
	Rascanya	0	-2,48491	2,80039	5041,7	9	19
	ALMASSERA	0	-1,79176	0,69921	3139,4	11	13
	LA POBLA DE FARNALS	0	-2,77259	0,51785	2180,5	10	5
	PUIG	0	-2,91777	0,87452	1980,7	9	6
HIPERCOR ADEMUZ	Extramurs	1	-1,50408	2,41902	6885,1	11	61
	Campanar	0	-1,09861	1,89388	6160,5	64	35
	La Saldia	0	-1,79176	2,34238	4209,7	17	29
	L'Olivereta	0	-1,60944	2,39285	12379,3	16	18
	Patraix	0	-2,07944	2,28922	5643,1	18	20
	Quatre Carreres	0	-2,35138	2,67762	7075,8	9	17
	Rascanya	0	-2,0149	2,24101	3791,7	12	24
	Benicalap	0	-1,79176	2,06348	4238,7	23	25
	Pobles de l'Oest	0	-1,8718	0,9583	5201	13	12
	BURJASSOT	0	-2,60269	1,99877	3602,2	10	13

HIPERMERCAT	MUNIC / DISTR	Primer	LN 1/T	LN %població	Competencia	PROCENQ	V. PREVISTOS
	PATERNA	0	-2,56495	2,207	4464	10	14
CONTINENTE EL SALER	L'Eixample	0	-1,38629	2,70946	5264,4	27	45
	Pla del Real	0	-1,94591	2,31318	4503,4	10	25
	Quatre Carreres	1	-0,91629	3,08522	3704,3	93	138
	Poblats Marítims	0	-2,60269	2,93323	2403,8	22	23
	Camins al Grau	0	-2,35138	2,73394	2969,8	44	25
	Algirós	0	-1,79176	2,57101	3479,6	30	35
	Pobles del Sud	0	-1,8718	1,7474	6031,6	9	17
PRYCA PATERNA	Benicalap	0	-2,19722	2,97928	5800,2	13	25
	Pobles de l'Oest	0	-1,79176	1,87409	6101	28	20
	PATERNA	1	-1,38629	3,1228	3506,4	105	109
	BURJASSOT	0	-1,94591	2,91456	3506,8	49	36
	MANISES	0	-1,79176	2,56671	3630,2	26	34
	GODELLA	0	-1,70475	1,63391	3906,7	15	22
	QUART DE POBLET	0	-2,07944	2,67512	4111,7	9	27
ALCAMPO PARQUESUR	Carabanchel	0	-2,35138	3,27372	3849,9	14	30
	Usera	0	-2,48491	0,31913	3651,5	13	6
	Villaverde	0	-2,07944	2,64928	3872,6	27	28
	LEGANÉS	1	-1,70475	2,96268	3324,2	125	85
	TOLEDO	0	-3,8712	1,90571	846,9	13	5
	FUENLABRADA	0	-2,60269	2,7935	3386	18	20
	GETAFE	0	-1,94591	2,75102	4420	21	31
ALCAMPO ALCORCÓN	Latina	0	-2,60269	3,59065	9809	18	18
	MÓSTOLES	1	-1,79176	3,24622	5615,5	110	77
	ALCORCÓN	0	-1,70475	2,9277	9595,6	57	27
	FUENLABRADA	0	-2,3979	2,96345	3625	35	25
PRYCA LAS ROZAS	Moncloa	1	-2,35138	3,6844	6553	12	59
	HOYO DE MANZANARES	0	-2,56495	0,25548	2046	12	6
	LAS ROZAS	0	-1,38629	2,57001	7370,4	99	36
	COLLADO VILLALBA	0	-2,63906	2,26873	1955,5	17	16

HIPERMERCAT	MUNIC / DISTR	Primer	LN 1/T	LN %població	Competencia	PROCENQ	V. PREVISTOS
	MAJADAHONDA	0	-1,79176	2,5245	7238,5	12	25
	TORRELODONES	0	-1,8718	0,96406	2902,9	31	15
	POZUELO DE ALARCÓN	0	-2,63906	2,87274	5144,3	11	17
	GALAPAGAR	0	-2,56495	1,21252	2046	12	10
EROSKI ALCOBENDAS	ALCOBENDAS	1	-1,09861	3,97404	9466,1	151	124
	SAN SEBASTIÁN DE LOS REYES	0	-1,09861	3,5942	7276,1	87	72
	TRES CANTOS	0	-1,60944	2,34319	2791,9	14	38
PRYCA HORTALEZA	Chamartín	0	-2,77259	3,10443	3903,7	13	19
	Tetuán	0	-2,89037	3,10696	3802,4	9	17
	Ciudad Lineal	0	-2,67415	3,53261	3649	18	27
	Hortaleza	1	-1,50408	3,05104	3240,4	224	101
HIPERCOR MÉNDEZ ÁLVARO	Arganzuela	0	-1,79176	2,42796	4041,1	40	31
	Retiro	0	-1,94591	2,51466	4259,3	40	28
	Carabanchel	0	-2,8622	3,11132	4232,4	16	17
	Puente de Vallecas	1	-1,79176	3,08271	4544,8	69	78
	Moratalaz	0	-2,60269	2,33082	4654,6	20	14
	Ciudad Lineal	0	-2,8622	3,07949	3730,1	15	18
CONTINENTE PLAZA ALUCHE	Latina	1	-0,69315	3,98281	5049,8	182	182
	Carabanchel	0	-2,80336	3,83584	4232,4	64	27
ALCAMPO PIO XII	Salamanca	0	-2,35138	2,58555	3206,5	9	22
	Chamartín	1	-1,25276	2,51602	2937,7	149	98
	Tetuán	0	-1,79176	2,51854	3519,1	12	34
	Chamberí	0	-2,89037	2,61139	3751	9	13
	Fuencarral	0	-3,04452	2,86737	3635,6	12	14
	Ciudad Lineal	0	-2,30259	2,9442	3810,1	47	27
	Hortaleza	0	-2,35138	2,46262	5053,1	21	18
ALCAMPO LA VAGUADA	Chamartín	0	-2,63906	3,3745	3744,8	19	25
	Tetuán	0	-2,07944	3,37703	3031,6	20	42
	Fuencarral	1	-1,94591	3,72585	2264,2	181	104
ALCAMPO MORATALAZ	Retiro	0	-1,60944	2,57911	3979,3	32	39

HIPERMERCAT	MUNIC / DISTR	Primer	LN 1/T	LN %població	Competencia	PROCENQ	V. PREVISTOS
	Salamanca	0	-2,48491	2,78528	2678,8	15	23
	Puente de Vallecas	0	-2,48491	3,14715	5928,1	10	21
	Moratalaz	0	-1,38629	2,39526	2841,7	139	46
	Ciudad Lineal	1	-1,94591	3,14393	2751,5	38	81
	San Blas	0	-2,70805	2,57829	2253,3	11	18
ALCAMPO VALLECAS	Retiro	0	-2,19722	3,17426	5481,6	16	29
	Puente de Vallecas	1	-1,09861	3,7423	4444,7	206	148
	Moratalaz	0	-2,19722	2,99041	4913,9	10	27
	Villa de Vallecas	0	-2,35138	2,41979	3434,7	23	20
	RIVAS-VACIAMADRID	0	-2,70805	1,01474	2442,6	9	7
ALCAMPO GETAFE	GETAFE	1	-1,60944	3,51431	4073,7	209	107
	PARLA	0	-2,07944	2,82914	4141	25	30
	TOLEDO	0	-3,7612	2,66901	853,2	20	8
	FUENLABRADA	0	-2,30259	3,55679	3574	14	37
PRYCA LOS ANGELES	Centro	0	-2,63906	2,90098	3414,8	9	20
	Arganzuela	0	-2,0149	2,73316	4899,1	9	28
	Carabanchel	0	-2,67415	3,41651	4135,9	14	24
	Usera	0	-2,07944	0,46192	3186,4	53	9
	Villaverde	1	-1,79176	2,79206	3405,4	145	74
	GETAFE	0	-2,48491	2,89381	4551,6	9	21
CONTINENTE MÓSTOLES	MÓSTOLES	1	-1,38629	3,66955	4732,2	199	123
	FUENLABRADA	0	-2,44235	3,38679	3647,1	9	30
	VILLAVICIOSA DE ODÓN	0	-1,60944	0,97922	3423,4	12	18
	ALCORCÓN	0	-1,60944	3,35104	9373,7	19	37
HIPERCOR ALCORCÓN	Latina	0	-2,19722	3,57336	8891,5	44	27
	ALCORCÓN	1	-0,91629	2,91042	5444,1	136	119
	FUENLABRADA	0	-2,56495	2,94617	3386	16	22
	MÓSTOLES	0	-2,07944	3,22893	2866,7	46	40
	VILLAVICIOSA DE ODÓN	0	-2,35138	0,5386	4042,1	12	7
HIPERCOR POZUELO	Moncloa	1	-1,70475	3,80167	3248,2	42	117

HIPERMERCAT	MUNIC / DISTR	Primer	LN 1/T	LN %població	Competencia	PROCENQ	V. PREVISTOS
	MAJADAHONDA	0	-2,07944	2,64177	5705,2	35	24
	POZUELO DE ALARCÓN	0	-2,30259	2,99001	3465,7	87	28
	LAS ROZAS	0	-1,94591	2,68727	6052,6	47	28
	BOADILLA DEL MONTE	0	-2,3979	1,88841	3292,9	16	15
PRYCA POZUELO	Moncloa	0	-2,25129	2,76357	6195,6	18	21
	Latina	1	-1,94591	3,69194	9121,2	147	66
	Carabanchel	0	-2,14007	3,54496	3722,6	26	41
	POZUELO DE ALARCÓN	0	-2,14007	1,95191	4489,2	42	17
	BOADILLA DEL MONTE	0	-2,3979	0,85031	4383,9	20	8
PRYCA MAJADAHONDA	Moncloa	0	-2,19722	3,80167	6594	10	36
	MAJADAHONDA	0	-0,69315	2,64177	5511,2	152	72
	LAS ROZAS	1	-0,69315	2,68727	6251,4	64	120
	BOADILLA DEL MONTE	0	-2,30259	1,88841	4704,1	16	14
	POZUELO DE ALARCÓN	0	-2,25129	2,99001	6349,5	16	23
PRYCA SAN FERNANDO	San Blas	0	-2,63906	3,20627	2505,3	24	25
	Barajas	0	-2,19722	1,9059	4190,2	24	16
	SAN FERNANDO DE HENARES	0	-1,70475	1,58628	3082,8	42	23
	COSLADA	0	-2,44235	2,64924	2458,1	61	23
	TORREJÓN DE ARDOZ	1	-1,79176	2,74403	5860,7	46	61
	PARACUELLOS DEL JARAMA	0	-2,8622	-0,17837	2799,7	14	3
	ALCALÁ DE HENARES	0	-2,80336	3,41983	1895,5	19	26
	MEJORADA DEL CAMPO	0	-2,48491	0,95869	2384	12	9
ALCAMPO TORREJÓN	TORREJÓN DE ARDOZ	1	-1,09861	3,13337	3094	136	132
	ALCALÁ DE HENARES	0	-2,60269	3,80916	1525,8	29	39
	SAN FERNANDO DE HENARES	0	-2,07944	1,97561	3286,2	14	20
	COSLADA	0	-2,60269	3,03857	2362,5	21	24
	MEJORADA DEL CAMPO	0	-2,67415	1,34802	2319,6	10	9
CONTINENTE PARLA	PINTO	0	-1,79176	2,68551	2768,5	18	39
	PARLA	1	-1,09861	3,83029	3241	167	160
	TOLEDO	0	-3,71357	3,67016	922,8	23	15

HIPERMERCAT	MUNIC / DISTR	Primer	LN 1/T	LN %població	Competencia	PROCENQ	V. PREVISTOS
CONTINENTE ALCOBENDAS	Chamartín	0	-2,70805	2,66257	3845,8	18	16
	Tetuán	0	-2,80336	2,66509	3814,2	12	15
	Fuencarral	0	-2,67415	3,01391	3161,8	35	21
	Ciudad Lineal	0	-2,60269	3,09075	3579,6	14	23
	Hortaleza	0	-2,63906	2,60917	4824,1	19	15
	ALCOBENDAS	1	-1,38629	2,03475	9697,9	64	45
	SAN SEBASTIÁN DE LOS REYES	0	-1,50408	1,65491	7808,1	37	19
	ALGETE	0	-2,67415	-0,03149	2297,1	15	4
	TRES CANTOS	0	-2,56495	0,4039	2547,4	16	6
	PRYCA ALCOBENDAS	Chamartín	0	-2,70805	3,3851	3832,9	9
Fuencarral		0	-2,67415	3,73645	3148,4	22	31
ALCOBENDAS		1	-0,69315	2,75729	6899,4	127	119
SAN SEBASTIÁN DE LOS REYES		0	-1,09861	2,37745	6542,8	47	44
ALGETE		0	-2,56495	0,69104	2196,2	9	7
ALCAMPO SANT QUIRZE	TERRASSA	0	-2,19722	3,46426	2844,5	17	41
	SANT QUIRZE DEL VALLÈS	0	-0,69315	0,60359	4411,3	30	30
	SABADELL	1	-1,38629	3,64515	3386,9	132	131
	BARBERÀ DEL VALLÈS	0	-1,60944	1,83998	5390,7	11	24
	RUBÍ	0	-2,48491	2,32136	2737,2	11	18
	CERDANYOLA DEL VALLÈS	0	-2,0149	2,43749	3918,8	18	26
CONTINENTE BARICENTRO	Nou Barris	0	-2,14007	3,19391	4499,8	24	33
	Sant Martí	0	-2,52573	3,32048	7954,6	12	19
	MONTCADA I REIXAC	0	-1,50408	1,22502	4113,1	14	21
	BARBERÀ DEL VALLÈS	0	-1,09861	1,39204	4784	44	31
	MOLLET DEL VALLÈS	0	-2,19722	1,66389	2917,7	20	16
	SABADELL	1	-2,07944	3,19721	5049,4	26	64
	RIPOLLET	0	-1,60944	1,24106	3861,9	25	20
	CERDANYOLA DEL VALLÈS	0	-1,60944	1,98955	3698,8	27	30
CONTINENTE MONTIGALÀ	Horta-Guinardó	0	-2,3979	2,8343	3006,3	10	25
	Nou Barris	1	-1,79176	2,85693	3919,6	12	74

HIPERMERCAT	MUNIC / DISTR	Primer	LN 1/T	LN %població	Competencia	PROCENQ	V. PREVISTOS
	Sant Andreu	0	-2,19722	2,59306	4665	10	23
	Sant Martí	0	-2,19722	2,98349	7580,8	13	22
	SANTA COLOMA DE GRAMENET	0	-1,60944	2,50773	3707,7	67	38
	BADALONA	0	-2,14007	3,00416	4544,9	143	30
EROSKI CORNELLÀ	SANT FELIU DE LLOBREGAT	0	-1,79176	1,90404	2517,3	8	27
	CORNELLÀ DE LLOBREGAT	1	-0,69315	2,74557	3576,5	164	140
	SANT JOAN DESPÍ	0	-1,25276	1,52173	3307,5	29	33
	SANT BOI DE LLOBREGAT	0	-1,79176	2,65961	6098,7	19	29
	L'HOSPITALET DE LLOBREGAT	0	-1,8718	3,9117	2674,5	31	65
	ESPLUGUES DE LLOBREGAT	0	-1,25276	2,18141	3381,7	14	45
PRYCA TERRASSA	TERRASSA	0	-1,79176	3,58092	2523,3	140	60
	SABADELL	1	-1,94591	3,76181	4783	25	91
	SANT CUGAT DEL VALLÈS	0	-2,35138	2,17987	2835,1	14	19
	VILADECAVALLS	0	-1,8718	-0,2586	2476,3	9	8
	RUBÍ	0	-2,07944	2,43801	2496,3	44	28
PRYCA PRAT	Eixample	0	-2,91777	3,18067	3598,2	15	18
	Sants-Montjuïc	0	-2,52573	2,75649	5526,8	24	17
	Les Corts	0	-2,74084	2,06354	6026,8	19	9
	SANT BOI DE LLOBREGAT	1	-0,91629	1,92235	3325,4	26	94
	CORNELLÀ DE LLOBREGAT	0	-1,8718	2,0083	6727,6	17	19
	L'HOSPITALET DE LLOBREGAT	0	-2,07944	3,17444	2825,4	68	39
	EL PRAT DE LLOBREGAT	0	-1,79176	1,7304	2212,6	38	25
	VILADECANS	0	-1,60944	1,44382	4077,5	23	22
	ESPLUGUES DE LLOBREGAT	0	-2,0149	1,44415	4758,8	9	14
PRYCA SANT ADRIÀ	Eixample	0	-2,60269	3,12233	3220	20	24
	Nou Barris	0	-2,35138	2,74876	4347,3	15	22
	Sant Andreu	0	-2,25129	2,4849	4454,9	20	21
	Sant Martí	0	-1,8718	2,87533	6794,7	82	29
	BADALONA	1	-1,79176	2,89599	3667,9	77	76
	SANT ADRIÀ DE BESÒS	0	-0,69315	1,03906	3261,4	34	41

HIPERMERCAT	MUNIC / DISTR	Primer	LN 1/T	LN %població	Competencia	PROCENQ	V. PREVISTOS
	SANTA COLOMA DE GRAMENET	0	-2,30259	2,39956	4390,2	21	19
CAPRABO BARNASUD	SANT BOI DE LLOBREGAT	0	-2,07944	3,69085	6640,4	20	37
	CASTELLDEFELS	0	-1,25276	2,83203	2738,2	52	63
	GAVÀ	0	-1,09861	2,89617	2998,1	95	71
	VILADECANS	1	-1,38629	3,21232	4047,5	47	109
MAXIM MARTORELL	ABRERA	0	-1,25276	2,26202	1556,1	31	53
	MARTORELL	1	-0,69315	3,37643	1902,1	152	178
	SANT ESTEVE SESROVIRE	0	-2,07944	1,53538	1250,1	18	19
	ESPARREGUERA	0	-1,79176	3,09849	1378,7	23	53
	OLESA DE MONTSERRAT	0	-2,3979	3,26935	1123	9	36
	GELIDA	0	-1,94591	1,92687	1371,5	13	26
	CASTELLVÍ DE ROSANES	0	-2,0149	0,14984	1371,5	9	10
CAPRABO SANT FELIU	SANT FELIU DE LLOBREGAT	1	-0,91629	3,21653	3022,3	163	148
	MOLINS DE REI	0	-1,38629	2,49771	2891,7	20	49
	SANT JUST DESVERN	0	-1,38629	2,13967	4347,2	16	36
	SANT JOAN DESPÍ	0	-1,60944	2,83422	5583,9	19	39
	ESPLUGUES DE LLOBREGAT	0	-1,70475	3,4939	5710,8	9	49
	VALLIRANA	0	-2,30259	1,49878	2127,4	15	14
HIPERCOR MERIDIANA	Nou Barris	0	-2,07944	3,53943	4601,3	41	41
	Sant Andreu	1	-1,38629	3,27556	3952,8	191	112
	Sant Martí	0	-2,0149	3,66599	7685,2	25	36
CONTINENTE GLÒRIES	Eixample	0	-1,94591	3,76802	2612,6	6	58
	Sant Andreu	0	-2,3979	3,13058	4757,4	8	25
	Sant Martí	1	-0,91629	3,52101	4459	240	151
CAPRABO L'ILLA	Eixample	0	-2,0149	3,10996	3498,2	63	38
	Sants-Montjuïc	1	-1,09861	2,68578	4914,8	39	102
	Les Corts	0	-0,91629	1,99283	3950,3	87	51
	Sarrià-Sant Gervasi	0	-2,0149	2,49461	3124,8	29	29
	Gràcia	0	-2,0149	2,35257	2919,9	15	27
	Nou Barris	0	-3,3322	2,73639	4537,2	19	9

HIPERMERCAT	MUNIC / DISTR	Primer	LN 1/T	LN %població	Competencia	PROCENQ	V. PREVISTOS
	Sant Martí	0	-2,77259	2,86295	7954,6	19	12
ESCLAT BARCELONA	Eixample	0	-2,19722	3,68482	3463,7	88	43
	Sants-Montjuïc	1	-0,91629	3,26064	4130,8	130	143
	Les Corts	0	-0,91629	2,56769	3452,5	26	68
	Sant Andreu	0	-3,17805	3,04738	4981	9	12

Cal recordar que PROCENQ ≡ Valors observats