A Multivariate Regression Model for the selection of locations for Commercial Establishments.

M.A. Mauricio Carvallo Ponton

This article describes the development of an statistical model applied to a decision-making process; discriminate among several location offers for commercial establishments. The reasons for developing this model were the fact that the relation between offers and locations accepted exceeds ten to one, and the national economic uncertainty threats all ill-defined economic feasibility projects. The main research was carried out not at the model development, but at the quantification of the different variables involved in it.

The model described in this article was developed in the biggest Commercial Group in Mexico, which has around a hundred stores of four diferents types, with their own price strategy, administration, locations, merchandise, etc. (i.e. the research was carried out at corporate level). The stores are all located throughout Mexico and cover a great range of socio-economic situations as well as traffic, topographic and competition conditions.

Before the model was implemented the decision-making process was the following:

a) An offer was made to the company by a third party.

b) If the plot of land was located in a city of interest to the company, the Department in charge visited it, otherwise the offer was declined.

c) After visiting the location the decision whether to carry on a field study or not was taken.

d) The field study was ill-defined, badly structered, and each time different to the previous one. Sometimes the economic potential was calculated from direct interviews to the neighbors. Other ocations, from statistical sources wich corresponded most of the time to greater extend of land, (delegacion, municipio, city, state).

e) Some other actors were included in the study, such as; competition, viality, traffic, etc.

f) From the study, monthly average sales was estimated by several pearsons (personal judgement), and a ROI was calculated.

g) The final decision was taken in The Executive Committe.

The process described lacked of a systematic approach to the problem. It was also a costly methodology due to several factors:

- The design , tranning, execution, processing and analysis of the field study.
- The "shadow price" of a wrong, late decision.
- The time required of several top managers of the Company to "estimate" the monthly average sales of the location.

After analysing the situation, The Company decides to do some research about it in order to improve the methodology. After revising the relevant literature (see References), it was found that several different models have been developed and implemented, such as the gravity model and the analogical one, but the one most used is the regression one. This model intends to "explain" the performance of a store (sales) due to several socio-econom c-demografic factors. Once the models have been discussed, the last one was chosen.

Variables Identification

The following variables were first thought as having some influence in the performance of an establishemt:

1. Type of Store. As it was mentioned, there are four different types of stores.
2. Total number of families in the Influence Zone (IZ).
3. Socio-economic level distribution of families in the IZ.
4. Exhibition area of the store.
5. Competition.
6. Traffic.

A sample of 23 stores was selected. This sample was not random because it was desired to cover a wide range of alternatives such as:

- Type of Store (the four of them).
- Age of Store (from one to almost 30 years old).
- Average Sales (from 70 to 300 million pesos per month on average).
- Location (covering all Mexico City and some other cities).
- Socio-economic level of the IZ (from level A to level D).

It was not desired to carry on variance analysis among all this variables, but the presence of them. If it were desired, a design of experiments such as that would have required a sample of at least 300 stores!, more than all The Company has.

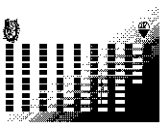
Field Study

First of all, the IZ of each store has to be defined. From 100 to 500 interviews were applied to random selected customers asking for their addresses. The IZ was determinated as that area covering from 50 to 75% of its customers. Some extrems were found as one store covering 97.5% of its customers in 2 Km. round, and another covering 35.9% in 3 Km. round.

Unfortunately, for all the stores of one specific type the IZ was competed spread. The merchandise of these stores in mainly fashion, expensive, nonperishable, and easy to transport. All this make the product easy to buy during lunch time in working days, or visiting friends during weekends. This press to do an special analysis for those stores, defining as 3 Km. round the IZ.

From different charts of the Delegaciones Politicas and Municipios actualiced to 1980, the blocks of each IZ were counted and numbered. A random sample fo 30 to 70 of them was chosen, and each block was visited and censed diferenciating each house/familie by socio-economic level A, B, C, D and E.

During the field study all competence stores were visited and their exhibition area was estimated, including the services departments.



Data Standardize

As Mexico is living with a high rate inflation, the dependant variable, monthly sales, was deflated to August 1983 using the National Price Index to Customers. In order to avoid seasonality, the average of the year was calculated.

After the IZ census of families, the field study end up with five figures for each store, too many to consider for a regression model with just 23 observations. Therefore they had to be standardized to a "socio-economic level equivalent". The main problem here was to determinate:

How many "socio-economic level X" is one "socio-economic level Y"?

Unfortunately there is no data about it except from The National Study of Income and Expenditure of 1977, where population was studied in decils (not in socio-economic levels), and some data of Marketing Services where an estimate of the distribution of socio-economic levels is presented in percentage. Another problem was to determinate the income and expenditure prices indexes for each socio-economic level.

Based in both studies, and estimate was calculated for gross income and expenditure for each socio-economic level. Then all the data was standardized to "socio-economic level E equivalent", based on net income.

Regression Models

After the data was standardized, it end up with seven figures for each store:

- Y Average Monthly sales, deflated to August 1983.
- X1 Type of store.
- X2 Total number of families of "socio-economic level E equivalent" in the IZ.
- X3 Exhibition area of competence.
- X4 Exhibition aea the store.
- X5 Percentage of customers comming from IZ.
- X6 Total number of families in IZ.

The build-up approach was followed incorporating the variables X1, X2, X3 and X4. An unexpected result was that the exhibition area of the store (X4) was not significant, (table I shows the analisys of significance). A similar analisys for the store of the specific type of spread IZ showed the same results.

Regression Model with 3 independant variables.

Correlation index 0.874.

Concept	D.F.	SS	MSS	F	level
Regression	3	9.52E10	3.17E10	31.38	99%
Error	13	1.36E10	1.01E09	–	–
Total	16	10.88E10	–	–	–

Regression Model with 4 independant variables. Correlation index 0.887

Concept	D.F.	SS	MSS	F	level
Due to X1 X2 X3	3	9.52E10	3.17E10	30.93	99%
X4	1	1.38E09	1.38E09	1.34	N.S.
Regression	4	9.65E10	2.42E10	23.54	99%
Error	12	1.23E10	1.03E09	–	–
Total	16	10.88E10	–	–	–

Table I Analisys of Significance

This result was welcome because in order to estimate the sales of a location, it is not needed to calculate potential the exhibition area of the store, a time-consumming operation.

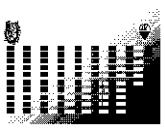
The accuracy of the model was satisfactory for two types of stores (2-4% error), acceptable for another (5-8% error), and unacceptable for the last one (10-25% error). The problem was that the sample for this type of stores was small, the stores were the oldest with, second and decoration, services, equipment, etc. Nevertheless this was considered a minor problem because grow for this type of stores is considered limited.

Sensitivity analisys were carried out to the inclusion of more variables of changing some of them, for instance: Dividing the total number of families of socio-economic level E equivalent by the percentage of customers comming from the IZ; Calculating the equivalence between socio-economic levels based on outcome instead of income; Dividing the exhibition area of competence by the area of the IZ; etc. None of them improved the results obtained with just 3 independant variables.

Conclusion

The developing of the model took three months of field study and one week of statistical analisys. It costed less than any previous single project had because it was developed in-house. Unfortunately, the proof of the model will take longer. However, thanks to a simple mathematical model a quick sales estimate can be obtained bases on a three-day field study, as a first filter.

It is important to mention that the model is not the decision, but an auxiliary tool for the decision-making process. There will always exist factors that may bias the decision (political strategy, Company policy, Competence strategy, etc.).



De los anuncios

Para anuncios en blanco y negro, el anunciante enviará negativo y prueba del mismo.

Para anuncios en negro y un color, negro y dos colores o cuatro colores (selección) se solicita el envío de negativo y juego de pruebas progresivas, asegurando que las medidas sean las indicadas al tamaño seleccionado.

Las fechas de cierre de los anuncios son 30 días previos a la aparición de cada número: 31 de Marzo, 30 de Junio, 30 de Septiembre y 31 de Diciembre.

Mayor información a los Tels. 547-77-92 y 541-62-07 o directamente en Carpio 471, Edif. D 3er Piso col. Sto. Tomás Deleg. Miguel Hidalgo México, D.F., atención Lic. José Antonio González R.

De las consideraciones de pago:

La presentación de facturas se hará al publicarse el anuncio.

Se otorgará un 15% de descuento por pronto pago (30 días).

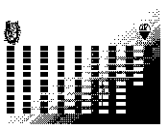
Las cancelaciones serán aceptadas si se presentan antes de las fechas de cierre para cada número.

Cualquier cambio en el material entregado por el cliente se facturará como Servicio Adicional.

Al ordenarse originales se facturará por la elaboración de los mismos, previa autorización del cliente.

No se acepta responsabilidad alguna por errores en los números clave de identificación en los anuncios.

El anunciante y/o su agencia de publicidad indemnizará y liberará a Investigación Administrativa de cualquier reclamación, queja o cargo de nivelación de derechos reservados, plagio, derecho de prioridad o cualquier otro cargo basado en el contenido o concepto de los anuncios.



Tarifa de los anuncios por ejemplar de aparición

4 tintas	1 - 2 números	3 - 4 números
4a. de forros	35,000.00	40,000.00
Blanco y negro		
3a. de forros	28,000.00	30,000.00
página interior	23,000.000	25,000.00
1/2 página horizontal	13,000.00	15,000.00
1 columna	13,000.00	15,000.00
1/2 columna	8,000.00	10,000.00

Las tarifas están sujetas a las fluctuaciones del mercado editorial.