

**GENERACION DE VIAJES EN EL GRAN SANTIAGO
Y UN CONTRASTE CON LAS PREDICCIONES DE ESTR AUS**

**Alan Thomas T.
Secretaría Ejecutiva
Comisión de Planificación de Inversiones en
Infraestructura de Transporte
Ahumada 48, piso 5
FAX: (562) 6966477**

**Rodrigo Parra G.
ICR Consultores Ltda.
Huelén 10, piso 6
FAX: (562) 2359758**

RESUMEN

Los modelos de generación de viajes han sido tradicionalmente de dos tipos: regresión lineal múltiple y análisis por categorías o de clasificación cruzada. Siguiendo esta última línea, en este estudio se analiza el método del análisis de varianza en dos factores (ANOVA) con efectos fijos y con un número arbitrario de observaciones por casilla, orientado a la selección de variables explicativas y a la detección de sus posibles interacciones, y el método de análisis de clasificación múltiple (MCA), orientado a la determinación de las tasas de generación de viajes por hogar.

Con los datos provenientes de la encuesta origen-destino de viajes del gran Santiago realizada en 1991 (EOD'91), se realizó un experimento en el que se consideró cuatro posibles variables explicativas para la tasa de generación de viajes diaria por hogar, determinándose sus estadígrafos y test de hipótesis.

Finalmente, las tasas de generación de viajes por hogar obtenidas utilizando el método MCA son contrastadas con las tasas predichas por los modelos de generación de viajes calibrados como parte del Estudio de Evaluación y Desarrollo del Sistema de Transporte Urbano de la Ciudad de Santiago (ESTRAUS). Para realizar lo anterior, los modelos provenientes del estudio ESTRAUS han sido alimentados con variables explicativas provenientes de la EOD'91.

1. INTRODUCCION

En este estudio se analiza el problema de la generación de viajes bajo la perspectiva del análisis por categorías (clasificación cruzada). En particular, nos interesará describir el análisis de varianza (ANOVA) en dos factores aplicado a tasas de generación de viajes. Como resultado del ANOVA se puede seleccionar en forma objetiva un conjunto de variables explicativas y sus respectivas estratificaciones. En este trabajo se presentan las expresiones correspondientes a las distintas varianzas y estadígrafos que surgen como consecuencia del ANOVA realizado en uno y dos factores.

Una vez seleccionadas las variables explicativas, es posible aplicar el método del análisis de clasificación múltiple (MCA), que a diferencia del análisis tradicional por categorías entrega tasas de generación de viajes confiables aún en las celdas correspondientes a los extremos de las categorías, las que seguramente verificarán un número insuficiente de observaciones.

A partir de la información obtenida de la Encuesta Origen-Destino de Viajes del Gran Santiago de 1991 (EOD'91), la que se aplicó en 31,000 hogares pertenecientes a las 34 comunas del Gran Santiago con un tamaño muestral del 3.125% (1), se obtienen mediante MCA, tasas de generación de viajes diarias utilizando distintas variables explicativas. En el marco del presente estudio hemos considerado adecuado explorar las siguientes variables para definir las categorías: ingreso familiar del hogar (INGRESO), número de automóviles por hogar (NAUTOS), tamaño o número de residentes por hogar (NRESID) y sector de residencia (SECTOR). Sobre cada una de las variables anteriores se determinan los estadígrafos que permiten discernir si la variable en cuestión aporta o no información en lo que a tasas de generación de viajes respecta, y si existen o no efectos interactivos entre las diversas variables explicativas.

Finalmente se utilizan los modelos de generación de viajes de ESTR AUS en fase predictiva utilizando las variables explicativas provenientes de la EOD'91, con la finalidad de contrastar las tasas de generación así obtenidas con las que provienen del MCA.

Cabe hacer notar que el método MCA ha sido escogido en el proceso de recalibración de los modelos de ESTR AUS, estudio iniciado en el presente año.

(1): En este estudio se utilizará el punto para denotar decimales y la coma para denotar miles.

2. ANALISIS DE VARIANZA Y DE CLASIFICACION MULTIPLE

En esta sección consideraremos el análisis de varianza (ANOVA) con efectos fijos y con un número arbitrario de observaciones por casilla. En primer lugar, nos referiremos al ANOVA en 1 factor y, a continuación, al ANOVA en 2 factores. El término efectos fijos dice relación con que siempre se consideran los mismos niveles de estratificación, es decir, los estratos se mantienen fijos durante los experimentos.

2.1. ANOVA en 1 Factor

El ANOVA en 1 factor con un número arbitrario de observaciones por casilla aplicado a generación de viajes ha sido tratado por Stopher (1975). Corresponde a estratificar los hogares por una única variable explicativa (por ejemplo, estrato de ingreso), designaremos por n_i a la cantidad de observaciones u hogares pertenecientes al i -ésimo estrato. La variable explicativa consta de I estratos, es decir, $i=1, \dots, I$. El número total de hogares lo denotamos por N .

La variable explicada corresponderá a una variable del tipo real tal como la tasa de generación de viajes diaria por hogar. Utilizaremos la letra x para denotar a la variable explicada.

A cada hogar perteneciente a la muestra encuestada en la EOD'91, se le determina su variable explicada x , lo anterior se realiza considerando los factores de corrección y expansión definidos en la EOD'91. Como cada hogar pertenece a un único estrato i , podemos denotar a la variable explicada de un hogar específico por x_{ij} con $j=1, \dots, n_i$.

La tasa de generación de viajes tendrá un valor promedio $\langle x_i \rangle$ al interior del estrato o casilla i .

El ANOVA en 1 factor consiste en poner a prueba la hipótesis nula (H_0) de que la tasa de generación de viajes no variará entre categorías. Para aceptar o rechazar H_0 , se descompone la variación total (SSt) existente en la muestra de hogares en 2 variaciones: una variación existente al interior de las categorías, denominada variación intra-categorías (SSw) y una variación explicada por la categorización, denominada variación inter-categorías (SSa).

$$SSt = \sum_{i=1}^I \sum_{j=1}^{n_i} (x_{ij} - \langle x_i \rangle)^2 + \sum_{i=1}^I n_i (\langle x_i \rangle - \langle x \rangle)^2 \quad (1)$$

En la expresión (1), $\langle x \rangle$ denota a la tasa promedio global. La suma de cuadrados inter-categorías puede ser descompuesta en dos términos, dando lugar a la siguiente expresión.

$$SSa = \sum_{i=1}^I \frac{\left(\sum_{j=1}^{n_i} x_{ij} \right)^2}{n_i} - \frac{\left(\sum_{i=1}^I \sum_{j=1}^{n_i} x_{ij} \right)^2}{N} \quad (2)$$

La variación total corresponderá a la suma de SSa y SSw, designando por SST a la variación total, ésta tendrá la siguiente expresión.

$$SSt = \sum_{i=1}^I \sum_{j=1}^{n_i} (x_{ij} - \langle x \rangle)^2 = SSa + SSw \quad (3)$$

A cada cantidad resultante de la suma total de cuadrados (SSt, SSa y SSw) se le asocia un valor entero denominado grados de libertad. A partir de la expresión (1), es sencillo ver que SSa tendrá (I-1) grados de libertad, lo anterior se debe a que las medias de los grupos se encuentran vinculadas a la media total, por lo que la variación del último estrato quedará determinada por las variaciones de los (I-1) estratos anteriores.

Del mismo modo, es posible deducir que SSw tendrá (N-I) grados de libertad, ya que las diferencias con la tasa promedio del estrato quitan un grado de libertad por cada categoría existente.

Finalmente, a partir de (3), se aprecia que al tomar las diferencias con respecto a la tasa promedio global, se perderá un grado de libertad, y como disponemos de N observaciones, los grados de libertad resultantes serán (N-1). Por lo tanto, se cumple que los grados de libertad correspondientes a SSt equivalen a los grados de libertad de SSa más los grados de libertad de SSw.

El siguiente paso es definir la media de cuadrados (MS) como la suma de cuadrados dividida por sus grados de libertad. Por lo tanto, MSa y MSw tendrán las siguientes expresiones.

$$MSa = \frac{SSa}{I-1} ; \quad MSw = \frac{SSw}{N-I} \quad (4)$$

Es posible demostrar que MSa y MSw son variables distribuidas chi-cuadrado (ver por ejemplo Glass y Stanley, 1986), por lo que un estadígrafo F puede ser formado tomando el cociente entre ambas variables ($F_a = MSa/MSw$).

Este valor es entonces comparado con el valor obtenido de las tablas de la distribución F con (I-1) grados de libertad para el numerador y (N-I) grados de libertad para el denominador. En nuestro caso debido a que N es un número bastante grande (~30,000 hogares), se entra a las tablas F con infinitos grados de libertad para el denominador.

2.2. ANOVA en 2 Factores

El ANOVA en dos factores aplicado al caso de un número arbitrario de observaciones por casilla es de una complejidad notoriamente mayor al caso del ANOVA en 1 factor mostrado anteriormente.

Stopher y McDonald (1983) presentan los resultados relativos a una aplicación concreta, pero no efectúan un desarrollo teórico. Por otra parte, muchos textos de estadística analizan el ANOVA en 2 factores; pero sólo llegan al caso de un número constante de observaciones por casilla. Glass y Stanley (1986) atacan el problema efectuando una extensión al caso de frecuencias proporcionales en las casillas, en este estudio adoptaremos tal enfoque denominado análisis de medias no ponderadas aplicable a diseños no proporcionales.

En el ANOVA de 2 factores existen 4 fuentes de variación: el factor A con sub-índice i , categorizado en I estratos; el factor B con sub-índice j , categorizado en J estratos; la interacción entre A y B, y las combinaciones dentro de las casillas de los niveles A y B.

Es posible establecer un modelo para los datos aplicable al ANOVA en dos factores, tal modelo tendrá los siguientes términos: un término μ , igual para todo dato, que describe la magnitud general de las observaciones; I términos α_i , uno por cada nivel del factor A; J términos β_j , uno por cada nivel del factor B; IJ términos $\alpha\beta_{ij}$, conocidos como términos de interacción que no corresponden al producto de α por β , y finalmente un término e_{ijk} por cada observación, que contribuirá a la identificación de diferencias entre las observaciones y la suma de los demás términos del modelo. En este modelo de los datos, una observación dada x_{ijk} se representará por:

$$x_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijk} \quad (5)$$

En el ANOVA en dos factores se contrastan las 3 siguientes hipótesis nulas:

-La variable explicada x no varía entre las categorías i , equivale a afirmar que para todo $i = \{1, \dots, I\}$: $\alpha_i = 0$.

-La variable explicada x no varía entre las categorías j , equivale a afirmar que para todo $j = \{1, \dots, J\}$: $\beta_j = 0$.

-No existe interacción entre los factores A y B, equivale a postular que la gráfica de las medias de población da origen a rectas paralelas o que para todo i y para todo j : $\alpha\beta_{ij} = 0$.

Las variaciones o sumas de cuadrados de desviaciones (SS), se determinan ahora mediante las siguientes expresiones (ver Glass y Stanley, 1986):

$$SSa = \sum_{i=1}^I \frac{\left(\sum_{j=1}^J \langle x \rangle_{ij} \right)^2}{J} - \frac{\left(\sum_{i=1}^I \sum_{j=1}^J \langle x \rangle_{ij} \right)^2}{I J} \quad (6)$$

$$SSb = \sum_{j=1}^J \frac{\left(\sum_{i=1}^I \langle x \rangle_{ij} \right)^2}{I} - \frac{\left(\sum_{i=1}^I \sum_{j=1}^J \langle x \rangle_{ij} \right)^2}{I J} \quad (7)$$

$$SSab = \sum_{i=1}^I \sum_{j=1}^J \langle x \rangle_{ij}^2 - SSa - SSb - \frac{\left(\sum_{i=1}^I \sum_{j=1}^J \langle x \rangle_{ij} \right)^2}{I J} \quad (8)$$

$$SSw = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^{n_{ij}} x_{ijk}^2 - \sum_{i=1}^I \sum_{j=1}^J \frac{\left(\sum_{k=1}^{n_{ij}} x_{ijk} \right)^2}{n_{ij}} \quad (9)$$

En las expresiones anteriores un hogar perteneciente en forma simultánea a los estratos i y j se denota por x_{ijk} con $k=1, \dots, n_{ij}$; el número de hogares en la casilla i, j , se denota por n_{ij} ; y el valor promedio de la variable explicada se denota por: $\langle x \rangle_{ij}$. Los grados de libertad son: de SSa ($I-1$), de SSb ($J-1$), de $SSab$ ($(I-1)(J-1)$) y de SSw ($N-IJ$). Finalmente, es necesario definir un factor de corrección c de la siguiente forma:

$$c = \frac{\sum_{i=1}^I \sum_{j=1}^J \frac{1}{n_{ij}}}{I J} \quad (10)$$

Las medias de cuadrados (MS) son ahora:

$$MSa = \frac{SSa}{I-1}; \quad MSb = \frac{SSb}{J-1}; \quad MSab = \frac{SSab}{(I-1)(J-1)}; \quad MSw = c \frac{SSw}{N-IJ} \quad (11)$$

Los estadígrafos F son:

$$F_a = \frac{MS_a}{MS_w}; \quad F_b = \frac{MS_b}{MS_w}; \quad F_{ab} = \frac{MS_{ab}}{MS_w} \quad (12)$$

Con las expresiones (10), (11) y (12) es posible poner a prueba las 3 hipótesis nulas mencionadas anteriormente.

2.3. El Método de Clasificación Múltiple (MCA)

Aplicando el ANOVA en 1 y 2 factores, mostrado anteriormente, es posible determinar un conjunto adecuado de variables explicativas, se procede a continuación a determinar un valor predicho de la variable explicada x'_{ij} . Si, por ejemplo, se consideran 2 variables explicativas, la expresión tendrá la forma siguiente (Stopher y McDonald, 1983):

$$x'_{ij} = \langle x \rangle + (x_i - \langle x \rangle) + (x_j - \langle x \rangle) = x_i + x_j - \langle x \rangle \quad (13)$$

Las tasas de generación así determinadas no presentarán los problemas inherentes al análisis de categorías clásico (bajo número de observaciones obtenidas en las categorías extremas).

3. VARIABLES EXPLICATIVAS Y SUS CATEGORIAS

Al emplear un método de análisis de clasificación múltiple, se hace necesario definir las variables explicativas y sus respectivas clases o rangos. En el marco del presente estudio se consideraron las siguientes variables descriptivas del hogar: estrato de ingreso familiar (INGRESO), número de automóviles por hogar (NAUTOS), número de residentes por hogar (NRESID) y sector de residencia (SECTOR). A continuación se describe cada una de las variables explicativas.

-INGRESO: Esta variable se considera estratificada en 5 categorías las que corresponden a agregaciones de los estratos de ingresos definidos en la EOD'91, tal agregación es idéntica a la adoptada en el estudio denominado "Análisis y Recalibración de los Modelos de ESTRAUS" (ESTRAUS 1993). En la tabla 1 se muestran las estratificaciones de ingresos definidas en la EOD'91 y las correspondientes a la variable INGRESO.

En la última columna de la tabla 1 aparece la categoría de ingreso del ESTRAUS 1986, tal pareo es necesario ya que posteriormente se realizará un contraste con las predicciones entregadas por los modelos de generación de ESTRAUS.

Tabla 1
Rangos de la variable INGRESO

Código EOD'91	Rango de ingresos (Pesos Mayo 1991)	Variable INGRESO (ESTRAUS 1993)	Categ.de Ingreso del ESTRAUS 1986
1	0- 41,000	1	1
2	41,100- 72,500	2	2
3	72,600- 110,400		3
4	110,500- 172,500	3	4
5	172,600- 262,000		
6	262,100- 405,000		
7	405,100-1,000,000	4	6
8	Más de 1,000,000	5	
9	No contesta	—	—

Durante la etapa de generación de viajes del estudio ESTRAUS 1986 se definieron 16 categorías, que resultan de cruzar las variables nivel de ingreso y posesión de automóvil. La tabla 2 muestra las 16 categorías de ESTRAUS definidas en 1986, en el mismo aparece también la actualización mediante IPC de los rangos de ingreso desde Noviembre de 1977 a Mayo de 1991, la que permitió efectuar la correspondencia anteriormente mostrada en la última columna de la tabla 1.

Tabla 2
Definición de las categorías ESTRAUS 1986

Categoría de Ingresos	Rango de ingresos (Pesos Nov. 1977)	Rango de ingresos (Pesos Mayo 1991)	Posesión de auto		
			0	1	2+
1	0- 2,500	0- 41,780	1	2	
2	2,501- 5,000	41,781- 83,560	3	4	
3	5,001- 10,000	83,561-167,120	5	6	7
4	10,001- 15,000	167,121-250,680	8	9	10
5	15,001- 20,000	250,681-334,240	11	12	13
6	más de 20,000	más de 334,240	14	15	16

-NAUTOS: Corresponde al número de autos existentes por hogar. Obviamente para posibilitar la correspondencia con ESTRAUS 1986 se mantuvo la estratificación de esta variable en 3 rangos: hogares sin posesión de auto, hogares con posesión de 1 auto y hogares con posesión de 2 o más autos.

-NRESID: Corresponde al número de residentes por hogar, en el marco de este estudio se ha considerado adecuado estratificarla en los siguientes 4 rangos: hogares con 1 residente, hogares con 2 o 3 residentes, hogares con 4 o 5 residentes y hogares con más de 5 residentes.

-SECTOR: Corresponde al sector de residencia del hogar. Los sectores de residencia son los mismos que se definieron en la EOD'91, es decir, los 6 siguientes: Norte (N), Poniente (P), Oriente (O), Centro (C), Sur (S) y Sur-oriente (S-O).

4. APLICACION AL CASO DE SANTIAGO (EOD'91)

El objetivo de esta sección es reportar los resultados de la aplicación del ANOVA y el MCA al caso de Santiago. Sin embargo, primeramente, se efectuará un análisis con respecto a los hogares con ingresos generados (no reportados).

4.1. Análisis de los Hogares con Ingresos no Reportados

En la EOD'91 se encuestó a 31,267 hogares, de los cuales sólo 29,182 reportaron ingresos, es decir, 2,085 hogares fueron ubicados durante el desarrollo de la encuesta en la categoría 9 (ingreso no reportado) de la tabla 1. Posteriormente, durante el desarrollo de un determinado estudio, a los hogares con ingresos no reportados se les generó un ingreso utilizando un método de búsqueda de hogares con características similares. Debe tenerse presente que el método empleado en la generación de los ingresos no hizo uso de la información relativa a viajes.

Hemos considerado de interés poner a prueba dicha generación de ingresos contrastando las tasas de generación de viajes correspondientes a ambos conjuntos de hogares. Los resultados, que se detallan en la tabla 3, arrojaron que, a nuestro criterio, la generación de ingresos es inaceptable, por lo que los análisis relativos al resto de las variables los haremos considerando exclusivamente a los 29,182 hogares con ingresos reportados.

Observando la tabla 3, se aprecia que los promedios de las tasas de generación de viajes diarias (x_i) son sistemáticamente menores en el caso de los hogares con ingresos generados. Se pensó que lo anterior podía deberse a que los hogares con ingresos generados tuvieran un menor número de residentes por hogar, sin embargo, los valores medios de estas variables resultaron ser prácticamente idénticos.

La varianza intra-categorías (MSw) resulta ser superior en el caso de los hogares con ingresos generados. Además, el mismo grupo de hogares verifica una varianza inter-categorías (MSa) inferior.

Pese a que, como veremos más adelante, el test F del ANOVA en 1 factor es muy poco exigente, la probabilidad de aceptación de la hipótesis nula (H_0) tiene, en el caso de los hogares con ingresos generados, un valor superior a 0.001. En contraste, en el otro grupo, el valor de la probabilidad de aceptación de la hipótesis nula resulta ser mucho menor que 0.001.

Tabla 3
Tasas de generación de viajes diarias
Contraste entre ingresos generados y reportados (EOD'91)

INGRESOS GENERADOS			INGRESOS REPORTADOS		
CAT	n _i	x _i	CAT	n _i	x _i
1	365	6.44	1	7400	5.72
2	721	6.85	2	13170	7.48
3	746	7.12	3	7210	8.07
4	224	7.49	4	1218	9.48
5	29	10.39	5	184	11.51
Total	2085	6.99	Total	29182	7.29
SSa=	528.11		SSa=	32143.15	
SSw=	77903.18		SSw=	1070861.49	
MSa=	132.03		MSa=	8035.79	
MSw=	37.45		MSw=	36.70	
Fa=	3.53		Fa=	218.95	
Probabilidad de aceptar Ho 0.001 < p < 0.01			Probabilidad de aceptar Ho p << 0.001		

4.2. Resultados del ANOVA en 1 Factor

En la tabla 4, se muestran los resultados del ANOVA en 1 factor aplicado sobre las variables explicativas: NAUTOS, NRESID y SECTOR.

Tal como se mencionara anteriormente, el test ANOVA en 1 factor resulta ser, en general, poco exigente. Lo anterior se corrobora observando la tabla 4, en la que se aprecia que las probabilidades de aceptación de H_0 tienen valores inferiores a 0.001. Lo anterior se cumple para el caso de las cuatro variables explicativas analizadas en el presente estudio.

Sin embargo, el ANOVA en 1 factor presenta la utilidad de poder efectuar un ranking entre las variables explicativas, es decir, es posible ordenarlas de mayor a menor F (mayor a menor explicatividad de la categorización). En este contexto, la primera variable resulta ser NRESID, seguida de NAUTOS e INGRESO. El F correspondiente a la variable SECTOR presenta un valor muy inferior al del resto de las variables.

Tabla 4
Tasas de generación de viajes diarias
Resultados del ANOVA en 1 factor: NAUTOS, NRESID y SECTOR

VARIABLE: NAUTOS			VARIABLE: NRESID			VARIABLE: SECTOR		
CAT	ni	xi	CAT	ni	xi	CAT	ni	xi
1: 0	20665	6.94	1: 1	393	4.75	1: N	3761	7.77
2: 1	6714	7.56	2: 2,3	10224	6.33	2: P	5906	7.68
3: 2+	1803	10.23	3: 4,5	15114	7.49	3: O	4768	7.57
			4: >5	1218	9.56	4: C	1658	6.60
						5: S	7779	6.73
						6: S-O	5310	7.29
Total	29182	7.29	Total	29182	7.29	Total	29182	7.29
SSa=	18570.18		SSa=	30336.34		SSa=	5330.13	
SSw=	1084434.46		SSw=	1072668.31		SSw=	1097674.52	
MSa=	9285.09		MSa=	10112.11		MSa=	1066.03	
MSw=	37.16		MSw=	36.76		MSw=	37.62	
Fa=	249.84		Fa=	275.06		Fa=	28.33	
Probabilidad de aceptar Ho p << 0.001			Probabilidad de aceptar Ho p << 0.001			Probabilidad de aceptar Ho p < 0.001		

4.3. Resultados del ANOVA en 2 Factores

Los resultados correspondientes al ANOVA en dos factores se muestran en las tablas 5 y 6.

En primer lugar, es interesante notar que las razones F dependen fuertemente de con qué otra variable se esté efectuando el análisis. Por ejemplo el F de INGRESO al cruzarlo con SECTOR es de 16.22, valor que cae a 2.95 al cruzar INGRESO con NAUTOS.

El ANOVA en dos factores tiende a confirmar que la variable SECTOR presenta poco poder explicativo.

Respecto de los efectos interactivos, es posible mencionar lo siguiente:

-Existe un gran efecto de interacción entre INGRESO y NRESID.

-Existen efectos interactivos considerables entre INGRESO y SECTOR y entre NAUTOS y SECTOR.

-Existen efectos interactivos moderados entre INGRESO y NAUTOS.

-No existen efectos interactivos entre NAUTOS y NRESID y entre NRESID y SECTOR.

El ranking de variables expuesto en el caso del ANOVA con 1 variable se mantiene, es decir, en orden decreciente de explicatividad se ubican: NRESID, NAUTOS, INGRESO y SECTOR. De las 4 variables anteriores, sólo la última es claramente inaceptable. Sin embargo, no se recomienda categorizar por INGRESO, debido a que presenta una gran interacción con NRESID y una interacción moderada con NAUTOS. En contraste, las variables NRESID y NAUTOS no presentan entre ellas efectos interactivos. Por lo tanto, como conclusión final, se recomienda categorizar en base a NRESID y NAUTOS.

Tabla 5
Resultados del ANOVA en 2 factores: Varianzas y estadígrafos F

Variables	MSa	MSb	MSab	MSw	Fa	Fb	Fab
INGRESO-NAUTOS	2.27	14.48	1.61	0.77	2.95	18.81	2.10
INGRESO-NRESID	2.01	46.44	5.60	0.55	3.63	83.73	10.10
INGRESO-SECTOR	40.73	11.17	13.02	2.51	16.22	4.45	5.18
NAUTOS-NRESID	3.02	12.97	0.74	3.22	0.94	4.03	0.23
NAUTOS-SECTOR	32.38	2.84	0.79	0.13	240.87	21.11	5.89
NRESID-SECTOR	25.66	0.69	0.24	0.24	105.91	2.85	0.99

Tabla 6
Resultados del ANOVA en 2 factores: Contrastes de hipótesis

Variables	Probab. Ho (A)	Probab. Ho (B)	Probab. Ho (AB)
INGRESO-NAUTOS	$0.01 < p < 0.05$	$p < 0.001$	$0.01 < p < 0.05$
INGRESO-NRESID	$0.001 < p < 0.01$	$p < 0.001$	$p < 0.001$
INGRESO-SECTOR	$p < 0.001$	$p < 0.001$	$p < 0.001$
NAUTOS-NRESID	$p > 0.25$	$0.001 < p < 0.01$	$p > 0.25$
NAUTOS-SECTOR	$p < 0.001$	$p < 0.001$	$p < 0.001$
NRESID-SECTOR	$p < 0.001$	$0.01 < p < 0.05$	$p > 0.25$

Las hipótesis nulas significan:

Ho(A): Las tasas de generación de viajes son iguales entre las categorías definidas con la primera variable.

Ho(B): Las tasas de generación de viajes son iguales entre las categorías definidas con la segunda variable.

Ho(AB): No existen efectos interactivos entre las categorías definidas con ambas variables.

4.4. Aplicación del método MCA

En el punto anterior se concluyó que es conveniente categorizar en base a NRESID y NAUTOS, por lo tanto, se aplicará el método MCA a las categorías definidas en base a estas variables. Además, con la finalidad de contrastar resultados, se aplicará el MCA a dos variables con fuertes efectos interactivos tales como NRESID e INGRESO. En la tabla 7 aparece la clasificación cruzada en NRESID y NAUTOS.

Tabla 7
MCA aplicado a variables explicativas: NRESID y NAUTOS

NRESID	NAUTOS	n_{ij}	x_{ij}	x^*_{ij}
1: 1	1: 0	371	4.70	4.40
1: 1	2: 1	21	5.50	5.01
1: 1	3: 2+	1	5.04	7.69
2: 2,3	1: 0	8676	6.27	5.98
2: 2,3	2: 1	1355	6.59	6.60
2: 2,3	3: 2+	193	7.12	9.27
3: 4,5	1: 0	10083	7.36	7.14
3: 4,5	2: 1	4161	7.43	7.76
3: 4,5	3: 2+	870	9.27	10.43
4: >5	1: 0	1535	8.59	9.21
4: >5	2: 1	1177	9.17	9.83
4: >5	3: 2+	739	12.19	12.50
TOTAL		29182	7.29	

n_{ij} : Número de hogares en la casilla i, j .

x_{ij} : Tasa promedio de generación de viajes en la casilla i, j .

x^*_{ij} : Tasa corregida por MCA para la casilla i, j .

Observando la tabla anterior, se aprecia que debido a que se está aplicando el método MCA a las variables que, según el ANOVA, presentaban mayor poder explicativo, la tasa promedio de viajes: x_{ij} presenta un comportamiento relativamente similar al de la tasa corregida: x^*_{ij} . Sin embargo, el efecto de la corrección por MCA es aumentar las diferencias en los valores de la tasa de generación de viajes entre categorías.

Por último, mirando el cuadro anterior, queda claro que no existen efectos interactivos entre NRESID y NAUTOS debido a que la tasa de generación de viajes es creciente con ambas variables. Como contraste al caso anterior, tenemos la tabla 8 en la que aparece la clasificación cruzada correspondiente a NRESID e INGRESO.

Tabla 8
MCA aplicado a variables explicativas: NRESID e INGRESO

NRESID	INGRESO	n_{ij}	x_{ij}	x^*_{ij}
1	1	258	4.39	3.18
1	2	115	5.52	4.94
1	3	20	4.86	5.53
1	4	0	0.00	6.93
1	5	0	0.00	8.97
2	1	3999	5.51	4.76
2	2	4810	6.86	6.52
2	3	1299	6.90	7.11
2	4	111	6.14	8.52
2	5	5	2.26	10.55
3	1	2846	6.01	5.92
3	2	7148	7.76	7.68
3	3	4467	7.92	8.27
3	4	594	7.96	9.67
3	5	59	8.14	11.71
4	1	297	6.96	7.99
4	2	1097	8.56	9.75
4	3	1424	9.67	10.34
4	4	513	11.96	11.75
4	5	120	13.56	13.78
TOTAL		29182	7.29	

n_{ij} : Número de hogares en la casilla i, j .

x_{ij} : Tasa promedio de generación de viajes en la casilla i, j .

x^*_{ij} : Tasa corregida por MCA para la casilla i, j .

Quedan claros los efectos interactivos entre NRESID e INGRESO: cuando NRESID es bajo, la tasa de generación de viajes es decreciente con INGRESO; y cuando NRESID es alto, la tasa de generación de viajes es creciente con INGRESO.

5. CONTRASTE CON LOS MODELOS DE GENERACION UTILIZADOS EN ESTRAUS

En 1986 por encargo de la Secretaría Ejecutiva de la Comisión de Transporte Urbano (SECTU), se desarrolló el Estudio de Evaluación y Desarrollo del Sistema de Transporte Urbano de la Ciudad de Santiago (ESTRAUS), cuyo objetivo principal es la evaluación de los denominados proyectos estratégicos de transporte. En la fase de calibración de los modelos de demanda, ESTRAUS utiliza datos provenientes de la encuesta Origen-Destino de Viajes 1977, los cuales fueron completados y actualizados con una encuesta más reducida tomada en el año 1986.

La Encuesta Origen-Destino de Viajes correspondiente a 1977 fue aplicada en 17 comunas del Gran Santiago, con un tamaño muestral de 3.8% de la población total, lo que significó encuestar a 22,000 hogares.

La modelación es realizada en dos periodos del día: punta mañana (7:00-9:00 hrs.) y fuera de punta (todo el día menos 7:00-9:00 y 17:30-19:30 hrs.).

Los propósitos de viajes son: trabajo basado en el hogar, estudio basado en el hogar, otros basado en el hogar y no basados en el hogar. Para los propósitos trabajo y estudio basados en el hogar, los modelos fueron calibrados considerando hogares chicos (menos de tres personas) y hogares grandes (tres o más personas).

Existen modelos correspondientes a los propósitos trabajo y estudio y a los periodos punta y fuera de punta. El propósito trabajo basado en el hogar fué calibrado con un modelo lineal del siguiente tipo:

$$Y = a_0 + a_1 T_1 + a_2 T_2 + a_3 T_3 + a_4 T_4 + a_5 T_5 + a_6 E_1 + a_7 E_2 + a_8 E_3 + a_9 NREST + a_{10} NP + a_{11} DC$$

Los términos involucrados corresponden a:

Y: número de viajes generados por hogar, para cada categoría ESTRAUS 1986.

T₁: Empresarios o directivo grandes y medianos, profesional universitario sin puesto directivo de importancia, empresario o industrial pequeño y medianos comerciantes.

T₂: Empleados públicos y privados, con escasa o ninguna función directiva.

T₃: Pequeños comerciantes, técnicos o artesanos que poseen taller, trabajadores no especializados y con trabajo estable.

T₄: Obreros especializados o con funciones directivas, obreros no especializados y sin funciones directivas.

T₅: Personal de servicio doméstico.

E₁: Estudiantes primarios.

E₂: Estudiantes secundarios.

E₃: Estudiantes secundarios o post secundarios.

DC: Dueñas de casa que no trabajan.

NP: Número total de personas en el hogar.

NREST: Número de personas restantes (NREST = NP - (T₁+E₁) -DC)

El propósito estudio basado en el hogar tiene un modelo de la siguiente forma:

$$Y = a_0 + a_1 E_1 + a_2 D \cdot E_1 + a_3 E_2 + a_4 D \cdot E_2 + a_5 E_3 + a_6 D \cdot E_3$$

Los términos involucrados corresponden a:

Y: Número de viajes generados por hogar con propósito estudio.

D: Variable muda (D=1 para EOD'86; D=0 para EOD'77).

Los modelos del estudio ESTRAUS fueron alimentados con las variables explicativas provenientes de la EOD'91 y se obtuvieron las tasas de generación de viajes presentadas en la tabla 9.

Para realizar lo anterior se efectuó una enumeración muestral cubriendo todos los hogares presentes en la muestra de la EOD'91, a continuación se agregaron las tasas de generación de viajes según las categorías del ESTRAUS 1986 (ver tabla 2). Las tasas son contrastadas con las determinadas mediante MCA para los mismos propósitos y períodos.

Para contrastar los resultados, se utilizó el índice de correlación definido como:

$$r^2 = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (14)$$

En la expresión anterior, σ_{xy} es la covarianza entre las tasas de generación y σ_x y σ_y son las desviaciones estándar.

Se aprecia que sólo para los viajes HB al trabajo efectuados en períodos punta se verifica una correlación positiva entre las tasas de generación de viajes.

Tabla 9
CONTRASTE ENTRE MODELOS DE GENERACION DE ESTRAUS Y EOD'91
TASAS DE GENERACION DE VIAJES POR PROPOSITO Y PERIODO

ESTRATO DE INGRESOS Y TASA DE MOTORIZACION		VIAJES HB AL TRABAJO				VIAJES HB AL ESTUDIO			
		PUNTA		NO PUNTA		PUNTA		NO PUNTA	
		EOD'91	ESTR.	EOD'91	ESTR.	EOD'91	ESTR.	EOD'91	ESTR.
1	0	0.62	0.47	0.99	0.74	0.39	0.89	0.87	0.89
2	0	1.02	0.67	1.45	0.94	0.52	0.61	1.10	1.03
3	0	1.07	0.99	1.42	0.96	0.69	0.69	1.23	1.07
4	0	1.10	1.59	1.62	0.70	1.24	1.27	1.87	1.77
5	0	1.28	1.65	1.81	0.54	2.29	1.30	2.58	1.91
1	1	0.59	0.55	0.94	0.90	0.75	0.64	1.37	0.54
2	1	0.99	0.70	1.40	0.89	0.88	0.49	1.60	0.87
3	1	1.04	0.82	1.37	0.87	1.05	0.75	1.73	0.79
4	1	1.07	0.97	1.57	0.89	1.60	0.38	2.37	0.52
5	1	1.25	0.63	1.76	0.91	2.65	0.32	3.08	0.35
1	2+	0.87	0.83	1.35	1.02	1.35	0.79	1.97	0.69
2	2+	1.27	1.06	1.81	1.02	1.48	0.26	2.20	0.48
3	2+	1.32	1.18	1.78	1.28	1.65	1.58	2.33	1.92
4	2+	1.35	1.13	1.98	0.57	2.20	0.50	2.97	0.53
5	2+	1.53	1.12	2.17	0.60	3.25	0.51	3.68	0.54
Correlación		+0.620		-0.263		-0.098		-0.163	

6. CONCLUSIONES

Para la determinación de tasas de generación de viajes es posible cuantificar estadísticamente el aporte de diversas variables explicativas utilizando el análisis de varianza en 1 y 2 factores. Lo anterior permite seleccionar variables explicativas y sus estratificaciones. Adicionalmente es posible medir los efectos interactivos de las variables explicativas, siendo recomendable utilizar variables con efectos interactivos mínimos.

Una vez seleccionadas las variables explicativas con ANOVA, puede predecirse la variable explicada con MCA, la que no poseerá los típicos problemas presentes en el análisis de categorías clásico relativos a las pocas observaciones presentes en las categorías extremas.

La utilización del método ANOVA permitió descartar a los hogares que no reportaron ingresos en la EOD'91, debido a que estos hogares presentaron un comportamiento estadísticamente distinto al del resto.

De las variables explicativas analizadas, NRESID fué la que tuvo mayor poder explicativo (respecto de la variable explicada tasa de generación diaria). Le siguieron: NAUTOS, INGRESO y SECTOR. El aporte de esta última variable resultó ser claramente marginal. Las dos variables que más aportan a explicar la tasa de generación de viajes diaria por hogar son, al mismo tiempo, las que presentan menores efectos interactivos.

No se recomienda categorizar por INGRESO, debido a que presenta una gran interacción con NRESID y una interacción moderada con NAUTOS. En contraste, es posible utilizar las variables NRESID y NAUTOS, al no detectarse efectos interactivos importantes.

De los 4 modelos de regresión lineal múltiple calibrados como parte del ESTRAUS 1986 y utilizados en fase predictiva con las variables provenientes de la EOD'91, sólo el correspondiente a viajes HB al trabajo efectuados en el periodo punta, verificó una correlación positiva con las tasas de generación obtenidas mediante MCA (para el mismo propósito-periodo).

7. REFERENCIAS

Glass, G.V. and Stanley, J.C. (1986). Statistical Methods in Educational and Psychology. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Stopher, P.R. (1975). Goodness-of-Fit Measures for Probabilistic Travel Demand Models. Transportation, 4(1), 67-83.

Stopher, P.R. and McDonald, K.G. (1983). Trip Generation by Cross-Classification: An Alternative Methodology. Transportation Research Record, 944, 84-91.