

ANÁLISIS DE CORRESPONDENCIAS APLICADO A ENCUESTAS DE TRANSPORTE. CASO BICICLETA

Laura Albrieu
Alejandro G. Baruzzi
Federico A. Baruzzi

Maestría en Ciencias de la Ingeniería: Mención Transporte
Secretaría de Ciencia y Técnica
Universidad Nacional de Córdoba

RESUMEN

En procura de mejorar el sistema de transporte en la ciudad de Córdoba, Argentina, se hace necesario implementar políticas que intenten canalizar los desplazamientos en vehículos no motorizados, bicicleta, como medio de transporte alternativo. Con motivo de recoger datos de este medio se realizaron encuestas sobre transporte urbano en polos atractores. Para el procesamiento de variables categóricas se aplicó el análisis de correspondencias, no encontrándose antecedentes de la aplicación de esta técnica a encuestas. Permite representar gráficamente en un subespacio óptimo filas y columnas de una tabla de contingencia ponderados por sus respectivos pesos. Las observaciones y variables fueron graficadas en un biplot de forma tal que se pueden hacer interpretaciones sobre las relaciones conjuntas. A modo de ejemplo se consideraron, medios de transporte a reemplazar por bicicletas y mejoras que habría que implementar en este medio para que se diera tal migración. Se concluyó que un alto porcentaje de los usuarios de otros medios migrarían al modo bicicleta si se incrementaran la seguridad vial o los espacios exclusivos de desplazamiento, no así los usuarios de motocicletas.

ABSTRACT

In order to improve the transport system in Córdoba, it is necessary to install policies that favor the movement of non –motorized vehicles as an alternative transport media. To collect data a survey on different transport media was made at different attraction poles. To process some categorical variables, correspondence analysis was applied. No preceding data of the application of this technique has been found. It is an exploratory technique that allows graphical representation of a contingency table, pondered by their respective weights, in an optimal sub-space. The results were presented in a biplot, where observations and variables are shown in such a way that interpretations can be made on their different relations. As an example, two variables were considered, transport media to be replaced by bicycles and the improvements that should be done so that this migration could take place. The conclusion was that a high percent of the users of other media would shift to the bicycle mode if road security and exclusive movement areas could be increased. This was not the case for motor-cyclists.

1. INTRODUCCION

Estudios realizados en mayo de 2012 por la Comisión de elaboración del plan integral de movilidad de la Ciudad de Córdoba, dependiente de la secretaría de Transporte y Tránsito de la Municipalidad de Córdoba, concluyeron que para mejorar el sistema de transporte logrando una movilidad urbana segura, fluida y confiable se hace necesario implementar políticas que intenten alentar los desplazamientos en vehículos no motorizados como medio de transporte alternativo no contaminante.

Dentro de ellos la bicicleta es una opción viable, cuenta con ventajas ambientales, ahorro de energía, bajo costo de adquisición, mantenimiento y operación, mayor velocidad en vías con alto flujo, menor espacio de circulación y estacionamiento que un automóvil, es un transporte puerta a puerta y además es beneficiosa para la salud. Como contrapartida tiene dificultad para llevar carga y pasajeros, es sensible a las condiciones ambientales, implica mayor riesgo de accidentes, vulnerabilidad a la violencia y robos, produce conflicto al manejar en vías en mal estado o compartidas con vehículos circulando a altas velocidades o con grandes diferencias de masa y presenta altas resistencias ante topografías adversas.

En el caso puntual de la ciudad de Córdoba, se presentan algunos factores tales como, clima templado, radio máximo de la periferia al centro de 8 km con una mayor concentración de viajes dentro de los 4 km y muy especialmente una gran cantidad de población estudiantil joven (Universidades), que motorizan a incentivar políticas que favorezcan a este medio de transporte no motorizado.

Para conocer la opinión, hábitos, y uso que podrían hacer de la bicicleta usuarios y no usuarios y poder determinar las necesidades y demandas que plantean respecto a este medio de transporte se realizaron encuestas. En las mismas se recopiló información sobre las distintas variables para tratar de establecer alguna relación entre ellas, presentar la situación actual como así también definir juicios declarados por individuos acerca de cómo actuarían frente a diferentes situaciones hipotéticas.

Para su visualización se pueden utilizar gráficos de dispersión o bien, gráficos de barras, donde se puede observar el comportamiento de una modalidad respecto de las distintas variables, pero en ambos, las relaciones en otra dimensión son solamente implícitas. En este estudio se pretende aplicar una metodología que permita observar los resultados de las encuestas en un mismo gráfico, de tal forma que se puedan hacer interpretaciones sobre las relaciones conjuntas de las variables y observaciones consideradas. La técnica aplicada es el análisis de correspondencias (AC) y se utilizó el biplot (B) para la representación gráfica.

La técnica del AC constituye una herramienta muy importante para el análisis de datos categóricos de una tabla de contingencia donde las variables y observaciones se representan como puntos en un espacio de baja dimensionalidad, generalmente bidimensional. Se puede interpretar como una generalización de un diagrama de dispersión.

La metodología aplicada consistió en confeccionar una tabla de clasificación cruzada, de contingencia, y luego realizar un análisis multivariado de correspondencias, que nos permite representar en un mismo gráfico, biplot, las filas (observaciones) y columnas (variables) de esa tabla.

2. RELEVAMIENTO DE LA INFORMACIÓN

Los lugares donde se realizaron las encuestas mencionadas se seleccionaron tomando como base la Encuesta de Origen / Destino 2009, Movilidad en el Área Metropolitana de Córdoba, Proyecto de Transporte Urbano en Areas Metropolitanas, PTUMA. Fue realizada durante la segunda mitad del año de 2009, la ciudad se dividió en zonas y se determinó que la cantidad de viajes diarios es de 2.705.310.

Para el presente trabajo se consideraron solamente los viajes diarios realizados en los modos: bicicleta, a pie, vehículo particular y ómnibus urbano. Los medios a pie, vehículo particular y ómnibus urbano, se relevaron por ser los usuarios de estos medios de transporte potenciales biciurbanos.

Las principales conclusiones, atendiendo lo expresado en el párrafo anterior, que se pudieron obtener son:

- Los viajes diarios realizados en la ciudad presentan una configuración radial, desde los barrios hacia el centro. (82,59%)

- Gran concentración de viajes en la zona céntrica (Barrio Centro, Nueva Córdoba y Ciudad Universitaria). Los traslados por motivos educacionales y laborales representan el 35% de los viajes diarios.
- La mayoría de los viajes diarios se realizan en ómnibus (41%), en segundo lugar en vehículo particular (35%), por último en medios no motorizados (23%) a pie y en menor proporción en bicicleta.
- Los viajes en ómnibus marcan la configuración radial de los desplazamientos con un gran porcentaje de viajes con origen o destino en la zona céntrica. La mayoría son viajes largos (son muy pocos los viajes registrados entre zonas vecinas).
- Los viajes a pie se registran en toda la ciudad, pero la mayor concentración se observa en la Zona Céntrica, Nueva Córdoba y Ciudad Universitaria con origen y destino en los barrios: Centro, Güemes, Alberdi, Nueva Córdoba y Ciudad Universitaria (42%).
- Se registran viajes en bicicletas en zonas puntuales de la ciudad, no en toda la ciudad, en la Zona Céntrica, en Nueva Córdoba y Ciudad Universitaria se registran muy pocos (4%).

Se confeccionaron planos de origen-destino de cada medio que vinculaban a todas las zonas, colocando cantidad de viajes ida-vuelta. En función de este mapeo, de las conclusiones precedentes y considerando la potencialidad de la demanda del modo bicicleta, se seleccionaron los puntos para realizar las encuestas, a saber:

1. Pabellón Argentina de la Universidad Nacional de Córdoba. Ciudad Universitaria.
2. Facultad de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional de Córdoba. Ciudad Universitaria.
3. Centro de compras Patio Olmos, ubicado en la esquina de Avda. Velez Sarsfield y Bvd. San Juan, al sur del área central.
4. Plaza San Martín, ubicada en el microcentro de la Ciudad de Córdoba.
5. Mercado Norte, mercado ubicado en la periferia del área central, al norte de la misma.

A modo de ejemplo se presenta en algunas zonas de la ciudad un mapeo de viajes en bicicleta Figura 1 (a), en ómnibus Figura 1 (b) y la ubicación de los lugares seleccionados para realizar las encuestas Figura 1(c).

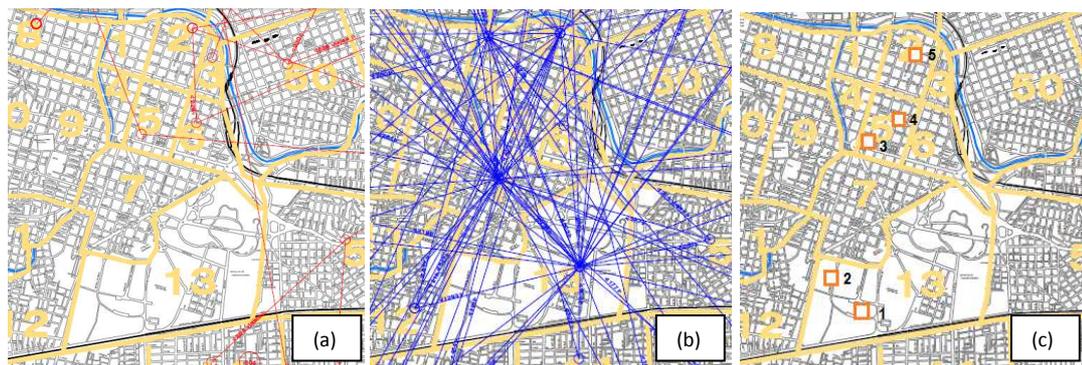


Figura 1. Mapeo de viajes en bicicleta (a), en ómnibus (b) y ubicación de los puntos donde se efectuaron las encuestas.

Durante los días 8,9,10 y 11 de Agosto de 2012 se realizaron las encuestas a usuarios y no usuarios de la bicicleta. En total se realizaron 247 encuestas, obteniéndose para cada variable consultada respuestas categóricas, en algunos casos nominales y en otros ordinales.

3. PROCESAMIENTO DE LA INFORMACION RELEVADA

Primero se confeccionó una planilla donde se condensó cada encuesta (observación) en una fila y en cada celda se colocó la respuesta correspondiente a cada variable. Ver tabla 1.

Tabla 1. Resultados de las encuestas.

FECHA	ENCUESTADOR	LUGAR DE ENCUESTA	Encuesta N°	SABE ANDAR EN BICI	SEXO	EDAD	NIVEL DE ESTUDIOS	OCCUPACION	FRECUENCIA DE USO DE BICI	CANTIDAD DE BICI PROPIAS	MOTIVO POR EL QUE USA BICI	CONDICIONES QUE DEBERIAN DARSE PARA QUE ANDUVIESE EN BICI	TIPO DE BICI	EN QUE EPOCA DEL AÑO UTILIZA LA BICI	EN QUE ASPECTO CREE QUE LA BICI LE AYUDA A MEJORAR LA SALUD	QUE CONDICIONES DEBERIAN DARSE PARA QUE SE DESPLAZASE EN BICI PARA TRABAJAR O ESTUDIAR ES SI REPARQUE CUAL MEDIO DE TRANSPORTE REEMPLAZARIA	QUE ESPACIO UTILIZA PARA EL DESPLAZAMIENTO	RELATIVO A LAS NORMAS DE CIRCULACION	MOTIVOS POR LOS CUALES INCREMENTARIA EL USO DE LA BICI	QUE SISTEMA DE TRANSPORTE EN BICICLETA UTILIZARIA SI EXISTIESE EL PUBLICO	
8-8	a	1	1	SI	MASC	A	S	E	MF	1	EJERC	TPO	OTROS	VAC	FGRAL	SC	VP	COM	C	EA	PUB
8-8	a	1	2	SI	MASC	A	T	E	MF	1	PASEO	ESPAC ADECI	MBIKE	IN	FGRAL	EE	AP	COM	NC	EA	PUB
8-8	a	1	3	SI	FEM	A	S	E	MF	1	NSABE	TPO	NO SABE	IN	FGRAL	SV	AP	COM	RS	EC	PUB
8-8	a	1	4	SI	FEM	A	T	E	MF	1	PLACER	TPO	NO SABE	IN	T	SV	VP	COM	C	CINT	PUB
-	-	-	-	SI	FEM	A	S	E	NU	1	PLACER	GUST	DE PASEO	NO USA	PP	EE	AP	COM	C	CEF	PUB
-	-	-	-	SI	MASC	A	S	E	NU	NING	NUSA	ESPAC ADECI	NOTGO	NO USA	FGRAL	EC	AP	NOUSA	C	EA	PRI
-	-	-	-	SI	FEM	A	S	E	M	1	PASEO	TPO	DE PASEO	IN	T	SV	TUP	COM	C	CEF	PUB
8-8	c	5	245	SI	MASC	A	S	E	NU	1	NUSA	NU	MBIKE	VAC	FGRAL	EE	TUP	C	C	EA	PUB
8-8	c	5	246	SI	FEM	A	S	E	MF	2	PASEO	TPO	DE PASEO	IN	FGRAL	SV	M	V	C	EA	PUB
8-8	c	5	247	SI	FEM	A	S	E	NU	NING	NUSA	POSEER	NOTGO	NO USA	T	EE	AP	NOUSA	NC	EA	PUB

Para cada polo se calcularon las frecuencias absolutas (Tabla 2) y relativas de cada variable (Tabla 3). Estos conjuntos de frecuencias relativas o vectores tienen características especiales debido a que la suma de sus elementos es 1 (100%).

Tabla 2. Frecuencias absolutas (a) y relativas (b) de los datos encuestados.

(a)												(b)													
LUGAR DE ENCUESTA	PATIO OLMOS PABELLON ARGENTINO MERCADO NORTE PLAZA SAN MARTIN FCEY y N	SABE ANDAR EN BICI		SEXO		EDAD		NIVEL ESTUDIO		QUE SISTEMA UTILIZARIA SI EXISTIESE EL PUBLICO		LUGAR DE ENCUESTA	PATIO OLMOS PABELLON ARGENTINO MERCADO NORTE PLAZA SAN MARTIN FCEY y N	SABE ANDAR EN BICI		SEXO		EDAD		NIVEL ESTUDIO		QUE SISTEMA UTILIZARIA SI EXISTIESE EL PUBLICO			
		SI	NO	MASC	FEM	A	B	S	T	PUB	PRI			SI	NO	MASC	FEM	A	B	S	T	PUB	PRI		
		49	1	24	26	46	4	35	15					19,84	0,40	9,72	10,53	18,62	1,62	14,17	6,07			15,79	4,05
		50	0	17	33	46	4	33	17					20,24	0,00	6,88	13,36	18,62	1,62	13,36	6,88			13,77	6,48
		50	0	14	36	47	3	39	11					20,24	0,00	5,67	14,57	19,03	1,21	15,79	4,45			16,60	3,24
		45	2	12	35	47	0	44	3					18,22	0,81	4,86	14,17	19,03	0,00	17,41	1,21			16,60	2,43
		49	1	36	14	41	9	42	8					19,84	0,40	14,57	5,67	16,60	3,64	17,00	3,24			12,55	7,69

Con las tablas anteriores se confeccionaron diagramas de barras que nos permitirían conocer las características dominantes de los encuestados como así también las preferencias ante determinadas condiciones. En la figura 2 (a) se observa que el 98,4% de los encuestados sabe usar la bicicleta, pero en la figura 2 (b) se observa que de esos usuarios sólo la usan para ir a trabajar o estudiar el 6,1%, o sea el 93,9% se desplaza hacia el centro o universidad o colegios en otros medios.

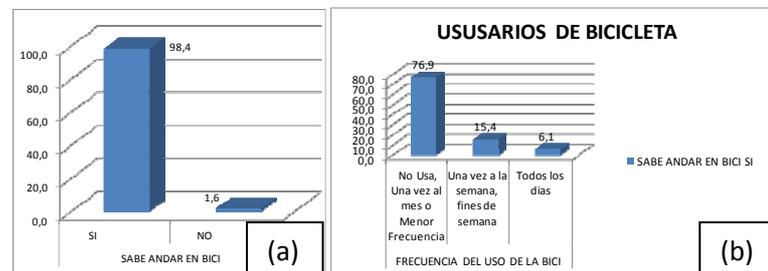


Figura 2. Porcentaje de encuestados que saben usar la bicicleta (a) y frecuencia de uso de bicicleta.(b)

En función de lo observado en figuras 2(a) y (b), se consultó sobre el medio de transporte utilizado y por las mejoras a implementar en el medio bicicleta para inducirlos a reemplazarlos. La Figura 3 nos grafica que medio de transporte tendría preferencia de

reemplazar por la bicicleta el encuestado en función de las mejoras que se implementaran en el desplazamiento en bicicleta pero analizadas independientemente, una por una, no se observa una correlación conjunta de todas las observaciones y variables.

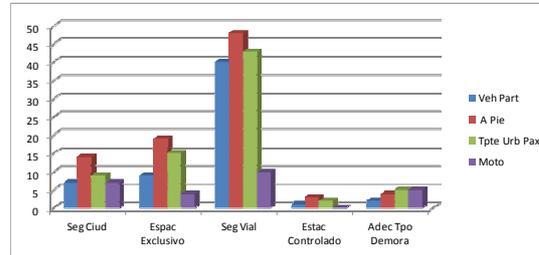


Figura 3. Medio de transporte a reemplazar por la bicicleta en función de las mejoras a implementar en este medio.

3.1. Análisis de correspondencias

Este punto fue realizado siguiendo la secuencia del libro La práctica del análisis de correspondencias del autor Michael Greenacre, primera edición, Julio 2008 y aplicando las técnicas incluídas en el software estadístico InfoStat, versión 2008.

Para la interpretación de los resultados de la encuesta se aplicó un análisis multivariado de correspondencias. El análisis de correspondencias (AC) constituye una herramienta de principal importancia para el análisis de datos textuales con los que se construyen tablas de contingencia, tabla a dos vías de variables categorizadas, relacionando el uso de varios vocablos entre distintos textos de discurso. Es una técnica exploratoria que permite representar gráficamente filas y columnas de dicha tabla de contingencia como puntos en un espacio Euclídeo de baja dimensión (generalmente bidimensional), se podría interpretar como una generalización de un Diagrama de Dispersión.

Confeccionada la tabla de clasificación cruzada (tabla de contingencia) se puede leer de dos formas distintas, como un conjunto de filas o de columnas. En cada caso ambas pueden ser conjuntos de frecuencias absolutas o relativas. A modo de ejemplo se presenta una sección de la tabla general de frecuencias relativas. Ver Tabla 3.

Tabla 3 Frecuencias relativas.

		LUGAR DE ENCUESTA					SABE ANDAR EN BICI		SEXO			QUE SISTEMA UTILIZARIA SI EXISTIESE EL PUBLICO		
		1	2	3	4	5	SI	NO	MASC	FEM	PUB	PRI	
LUGAR DE ENCUESTA	1	20,24	0,00	0,00	0,00	0,00	19,84	0,40	9,72	10,53				15,79	4,05
	2	0,00	20,24	0,00	0,00	0,00	20,24	0,00	6,88	13,36				13,77	6,48
	3	0,00	0,00	20,24	0,00	0,00	20,24	0,00	5,67	14,57				16,60	3,24
	4	0,00	0,00	0,00	19,03	0,00	18,22	0,81	4,86	14,17				16,60	2,43
	5	0,00	0,00	0,00	0,00	20,24	19,84	0,40	14,57	5,67				12,55	7,69
SABE ANDAR EN BICI	SI	19,84	20,24	20,24	18,22	19,84	98,38	0,00	41,70	56,68				74,09	23,48
	NO	0,40	0,00	0,00	0,81	0,40	0,00	1,62	0,00	1,62				1,21	0,40
SEXO	MASC	9,72	6,88	5,67	4,86	14,57	41,70	0,00	41,70	0,00				29,55	11,74
	FEM	10,53	13,36	14,57	14,17	5,67	56,68	1,62	0,00	58,30				45,75	12,15
.....														
														
QUE SISTEMA UTILIZARIA SI EXISTIESE EL PUBLICO	PUB	15,79	13,77	16,60	16,60	12,55	74,09	1,21	29,55	45,75				75,30	0,00
	PRI	4,05	6,48	3,24	2,43	7,69	23,48	0,40	11,74	12,15				0,00	23,89

Tomando como observaciones las filas en donde se encuestaba las condiciones primordiales que deberían darse para que se desplace en bicicleta (SC: seguridad ciudadana, EE: espacio

exclusivo, SV: seguridad vial, EC: estacionamiento controlado y ATD: adecuado tiempo de desplazamiento o viaje) y como variables el medio que tendría preferencia en reemplazar si se aseguraba cumplir con esa condición (VP: vehículo particular, AP: a pie, TUP: transporte urbano de pasajeros y M: moto) se puede inferir que si aseguramos seguridad vial y espacios exclusivos para circulación un 76,12% de los encuestados tendría preferencia en migrar al modo bicicleta desde los otros medios, un 14,98% se siente afectado por la seguridad ciudadana, un 6,48% considera el tiempo de traslado como prioritario y solo un 2,43% se ve afectado por la falta de un estacionamiento controlado en el lugar de destino. Lo anteriormente expuesto se observa en la tabla 4.

Tabla 4. Frecuencias relativas relacionando condiciones a mejorar en el medio bicicleta para que los usuarios de otros medios prefieran migrar a la bicicleta.

		SI SE CUMPLIERA ESA CONDICION QUE MEDIO DE TRANSPORTE ESTARIA DISPUESTO A REEMPLAZAR				TOTALES
		Veh Particular (VP)	A Pie (AP)	Tpte Urbano Público(TUP)	Moto (M)	
QUE CONDICIONES DEBERIAN DARSE PARA QUE SE DESPLACE EN BICI	Seguridad Ciudadana (SC)	2,83	5,67	3,64	2,83	14,98
	Espacio Exclusivo (EE)	3,64	7,69	6,07	1,62	19,03
	Seguridad Vial (SV)	16,19	19,43	17,41	4,05	57,09
	Estacionamiento Controlado (EC)	0,40	1,21	0,81	0,00	2,43
	Adecuado Tiempo Desplazamiento (ATD)	0,81	1,62	2,02	2,02	6,48

En la tabla 4 se observa por ejemplo que un 16,19% de los usuarios de VP podrían migrar a la bicicleta si aumentara la SV. Siguiendo en este ejemplo, de dos entradas con varias modalidades cada una, podemos calcular para cada fila, en cada celda, el porcentaje correspondiente de cada observación respecto al total de la fila, de la misma manera se puede proceder con las columnas o variables, obtenemos así los perfiles filas y columnas. En la Tabla 5 se observa las frecuencias relativas por fila, perfiles fila, y las frecuencias relativas por columnas, perfiles columnas.

Tabla 5. Perfiles fila, columna y totales.

Perfiles fila (frecuencias relativas por filas)						Perfiles columna (frecuencias relativas por columnas)					
En columnas: QUE MEDIO DE TPTE REEMPLAZARIA						En columnas: QUE MEDIO DE TPTE REEMPLAZARIA					
En filas: CONDICIONES PARA MIGRAR A BICICLETA						En filas: CONDICIONES PARA MIGRAR A BICICLETA					
	VP	AP	TUP	M	Total		VP	AP	TUP	M	Total
SC	18,92	37,84	24,32	18,92	100	SC	11,86	15,91	12,16	26,92	14,98
EE	19,15	40,43	31,91	8,51	100	EE	15,25	21,59	20,27	15,38	19,03
SV	28,37	34,04	30,5	7,09	100	SV	67,8	54,55	58,11	38,46	57,09
EC	16,67	50	33,33	0	100	EC	1,69	3,41	2,7	0	2,43
ATD	12,5	25	31,25	31,25	100	ATD	3,39	4,55	6,76	19,23	6,48
Total	23,89	35,63	29,96	10,53	100	Total	100	100	100	100	100

A cada punto fila se le puede asignar un peso (masa) que se obtiene a través de la división del total de fila por el total de la muestra. Podemos obtener el perfil fila medio como el centroide de los perfiles fila ponderando cada perfil con su correspondiente masa. Estos conceptos se pueden observar en la Tabla 6. Cuando se ponderen los puntos de cada perfil fila, el centroide de la representación gráfica no se situará en el centro geográfico de la nube de puntos sino que tiende a situarse cerca de los puntos con mayor masa. Si trabajamos con columnas el procedimiento es equivalente. El AC analiza de forma similar filas y columnas de una tabla de contingencia, o sea trata filas y columnas de forma simétrica.

Tabla 6. Tabla que muestra frecuencias absolutas, el perfil fila, el perfil fila medio y las masas de cada fila.

		SI LA RESPUESTA ANTERIOR ES SI, MARQUE CUAL MEDIO DE TRANSPORTE REEMPLAZARIA								Total	Masa de las filas
		VP		AP		TUP		M			
		FA	FR	FA	FR	FA	FR	FA	FR		
QUE	SC	7	18,92%	14	37,84%	9	24,32%	7	18,92%	37	0,150
CONDICIONES	EE	9	19,15%	19	40,43%	15	31,91%	4	8,51%	47	0,190
DEBERIAN DARSE	SV	40	28,37%	48	34,04%	43	30,50%	10	7,09%	141	0,571
PARA QUE SE	EC	1	16,67%	3	50,00%	2	33,33%	0	0,00%	6	0,024
DESPLACE EN BICI	ATD	2	12,50%	4	25,00%	5	31,25%	5	31,25%	16	0,065
	Total	59		88		74		26		247	1,000
PERFIL FILA MEDIO			23,89%		35,63%		29,96%		10,53%		

Podemos representar gráficamente los perfiles fila y columna en un espacio que tenga tantos vértices como variables, en el ejemplo tendría cuatro, o sea un tetraedro. Las filas de la tabla de contingencia pueden ser vistas como puntos con coordenadas dadas por las columnas de la tabla o sea dentro de ese espacio es posible ubicar los perfiles fila atendiendo las coordenadas que cada fila tiene respecto a cada columna. El grado de dispersión de los perfiles fila respecto del centroide fila nos da una idea sobre la variabilidad existente en la tabla de contingencia, cuanto más cerca se hallen los perfiles del centroide menor será la variabilidad. En la tabla 6 se puede observar que hay variabilidad de los valores de los perfiles respecto al perfil fila medio en las columnas VP, AP y M, en menor medida en TUP, lo que indicaría que ante la implementación de cualquier mejora prácticamente el 30% de los encuestados tendría la preferencia de migrar al modo bicicleta desde el TUP.

La homogeneidad de la tabla de contingencia se puede calcular, en general, con el concepto de chi-cuadrado. Este estadístico también se puede aplicar en AC, gráficamente, para calcular la distancia entre los perfiles (una interpretación geométrica del estadístico calculado a partir de los datos de la tabla de contingencia).

Si no existieran diferencias entre las condiciones que deberían darse para que el encuestado se desplace en bici en lo concerniente a los distintos medios de los que tendría preferencia a migrar, los perfiles de todas las filas deberían ser semejantes al perfil fila medio y el valor del estadístico sería muy bajo. Para obtenerlo, se calculan los perfiles esperados para cada fila en función de los valores del perfil fila medio. Las frecuencias esperadas generalmente difieren de las observadas, lo que se desea saber es si esas diferencias son suficientemente grandes como para contradecir la hipótesis de que las filas son homogéneas. Esa medida de discrepancia entre lo observado y lo esperado es el estadístico chi-cuadrado (X^2) Ecuación 1. Cuanto mayor sea este valor, mayores serán las discrepancias entre las frecuencias observadas y las esperadas, en consecuencia estaremos menos convencidos de la certeza del supuesto de homogeneidad.

$$X^2: \sum (VO - VE)^2 / VE \quad (1)$$

Donde: X^2 : chi-cuadrado
 VO : valor observado;
 VE : valor esperado.

En la Tabla 7 se presenta un resumen del aporte a chi-cuadrado por filas y columnas y el cálculo de la probabilidad de significación, p , calculado utilizando el programa Estadístico InfoStat.

Tabla 7. Contribuciones al chi-cuadrado por celda y cálculo del p

Contribuciones por celda al estadístico chi-cuadrado						Frecuencias absolutas					
En columnas: SI LA RESPUESTA ANTERIOR ES SI, QUE MEDIO DE TPTE REEMPLAZARIA						En columnas: SI LA RESPUESTA ANTERIOR ES SI, QUE MEDIO DE TPTE REEMPLAZARIA					
En filas: QUE CONDICIONES DEBERIAN DARSE PARA MIGRAR AL MODO BICICLETA						En filas: QUE CONDICIONES DEBERIAN DARSE PARA MIGRAR AL MODO BICICLETA					
	VP	AP	TUP	M	Total		AP	M	TUP	VP	Total
SC	0,38	0,05	0,39	2,48	3,3	ATD	4	5	5	2	16
EE	0,44	0,3	0,06	0,18	0,99	EC	3	0	2	1	6
SV	1,19	0,1	0,01	1,58	2,88	EE	19	4	15	9	47
EC	0,13	0,35	0,02	0,63	1,13	SC	14	7	9	7	37
ATD	0,87	0,51	0,01	6,53	7,91	SV	48	10	43	40	141
Total	3,01	1,31	0,5	11,4	16,21	Total	88	26	74	59	247
						Estadístico		Valor	gl	p	
						Chi Cuadrado Pearson		16,21	12	0,1817	
						Chi Cuadrado MV-G2		14,53	12	0,2679	

El valor p asociado de 0,1817 a un valor estadístico X^2 de 16,21 (Tabla 7) nos indica que la probabilidad de que las frecuencias observadas se correspondan con el supuesto de homogeneidad es extremadamente baja, o sea se rechaza la homogeneidad de la tabla de contingencia y concluimos que es muy probable que existan diferencias entre las variables y las observaciones. Esto confirma lo observado en la tabla 6.

Como se expresó anteriormente, al estadístico chi-cuadrado podemos aplicarlo en AC, gráficamente, para medir la distancia entre los perfiles. Si se deseara calcular la distancia euclídea de una observación, por ejemplo ATD, al perfil fila medio se aplicaría la ecuación 2.

$$Dist. eucl.: \sqrt{\sum((perf\ i\ esima\ fila\ y\ j\ esima\ columna) - (perf\ fila\ medio\ de\ la\ columna\ j))^2} \quad (2)$$

Dando valores :

$$Dist. eucl.: ((0,125-0,2389)^2 + (0,25-0,3563)^2 + (0,3125-0,2996)^2 + (0,3125-0,1053)^2)^{1/2}$$

Si deseáramos representar gráficamente en dos dimensiones las distancias X^2 entre el perfil fila y el centroide, encontraríamos que estas distancias no son las distancias euclídeas (sin ponderar) sino que se encuentran distorsionadas ya que están ponderadas por las masas, se produce una transformación de los valores de las distancias geométricas de los perfiles al centroide. Podemos calcular la distancia euclídea o pitagórica ponderada entre la fila i y el centroide o distancia chi-cuadrado aplicando la ecuación 3.

$$Dist. X^2: \sqrt{\sum \frac{((perf\ i\ esima\ fila\ y\ j\ esima\ columna) - (perf\ fila\ medio\ de\ la\ columna\ j))^2}{perf\ fila\ medio\ de\ la\ columna\ j\ o\ masa\ de\ la\ j\ esima\ columna}} \quad (3)$$

Dando valores:

$$Dist. X^2: ((0,125-0,2389)^2 / (0,2389) + (0,25-0,3563)^2 / (0,3563) + (0,3125-0,2996)^2 / (0,2996) + (0,3125-0,1053)^2 / (0,1053))^{1/2}$$

Otra forma de medir la variabilidad de las proporciones de la tabla de contingencia es a través del concepto de inercia de la tabla. Ver ecuación 4.

$$Inercia: x^2 / n \quad (4)$$

Donde x^2 : chi-cuadrado de la tabla

n : total de la muestra

$$Inercia: 16,21/247=0,066$$

O lo que es lo mismo podemos calcular la inercia aplicando la ecuación 5.

$$Inercia: \sum(i\ ésimo\ masa) \times (distancia\ x^2\ de\ i\ ésimo\ perfil\ media)^2 \quad (5)$$

Dado que la suma de las masas es 1, podemos decir que la inercia es la media ponderada de los cuadrados de las distancias X^2 entre los perfiles fila y su perfil media, geoméricamente mide lo lejos que se hallan los perfiles fila de su perfil medio, donde el perfil medio simboliza la hipótesis de homogeneidad de los perfiles.

3.1.1. Reducción de la dimensionalidad.

Obtenida la matriz de desviaciones Chi cuadrado, tabla 7, el AC opera sobre la misma, determinando (bajo la hipótesis de independendencia entre filas y columnas) un subespacio óptimo para la representación de los perfiles filas y columnas ponderados por sus respectivos pesos. Las observaciones multivariadas se grafican en planos para así poder identificar las asociaciones de mayor peso entre las modalidades de varias variables cualitativas.

Si analizamos la tabla 6, es una tabla multidimensional, tiene cuatro columnas (cuatro categorías de modos de transporte a reemplazar), conlleva perfiles columnas cuatridimensionales, estos cuatro puntos generan un cuerpo, tetraedro. Los perfiles fila son puntos que se sitúan en un espacio de cuatro dimensiones y como cada uno, por lo tanto, tiene cuatro elementos y su suma es igual a 1 éstos se sitúan en un espacio de una dimensión menos, o sea la tabla se puede representar en un espacio tridimensional. Los perfiles constituídos por m elementos se sitúan en espacios de dimensionalidad m-1. Definido el cuerpo (en este caso de cuatro columnas) debemos ubicar los perfiles fila en el mismo.

En la mayoría de las aplicaciones de AC las tablas de contingencia pueden tener más filas y columnas, en consecuencia, los perfiles se ubican en espacios de mayor dimensionalidad. Para visualizar la posición relativa de los perfiles es necesario reducir esa dimensionalidad pero eso implica una pérdida de información, ya que esas proyecciones no tienen sus verdaderas posiciones. La esencia del AC es la identificación de subespacios (S) de pocas dimensiones que contengan los perfiles, aunque sea de forma aproximada, trabaja con dimensiones para las cuales existe muy poca dispersión de los perfiles y elimina las direcciones de dispersión que aportan poca información. El criterio para seleccionar (S) es calcular la proximidad de todos los perfiles al mismo como la suma ponderada de distancias al cuadrado y minimizarla, por lo tanto (S) pasa por el centroide de los puntos. Ver ecuación 6.

$$\text{Proximidad a subespacios (S)}: \sum m_i [d_i(S)]^2 \quad (6)$$

Donde:

m_i : masa del i-ésimo perfil

$d_i(S)$: distancia X^2 entre el i-ésimo perfil de masa m_i y S.

Encontrado el (S) óptimo, hay que proyectar los perfiles sobre él y observar sus proyecciones como una aproximación a sus verdaderas posiciones en el espacio de mayor dimensionalidad. Para determinar la discrepancia entre las posiciones exactas y las aproximadas se utiliza la inercia total de los perfiles, que representa una medida de la variabilidad total, es decir una medida de la dispersión geométrica de los puntos de sus verdaderas posiciones. La precisión del ajuste se expresa como % de inercia. En la tabla 8 se muestra el valor de inercia y la contribución de X^2 para cada eje del plano seleccionado en este ejemplo.

Tabla 8. Contribución a la chi-cuadrado por eje.

Contribución a la Chi cuadrado					
	Autovalor	Inercias	Chi-Cuadrado	(%)	% acumulado
1	0,23	0,05	13,38	82,56	82,56
2	0,1	0,01	2,25	13,89	96,45

El eje 1 es la dimensión que mejor ajusta ya que explica la mayor inercia, el eje se denomina eje principal y la inercia explicada inercia principal. Si sumamos las inercias parciales de cada eje obtenemos 0,06, valor que coincide con el calculado en la ecuación 4. Las proyecciones de los cinco perfiles fila explican el 96,45% de la inercia total. Está claro que los cinco perfiles fila se hallan muy cerca del plano representado. También se pueden proyectar los vértices sobre el (S) y utilizarlos como puntos de referencia solamente, ya que se encuentran a una escala diferente.

3.1.2. Representaciones bidimensionales. Biplots.

Definido el subespacio (S), para verlo gráficamente se adopta una representación bidimensional, ya que si aumentamos la dimensionalidad del espacio si bien se gana en exactitud se pierde en visualización de los puntos.

Una posibilidad para la representación conjunta de filas (observaciones) y columnas (variables) son los biplots, que nos permiten hacer interpretaciones sobre las relaciones conjuntas. El prefijo bi refleja la característica de que tanto observaciones como variables son representadas en el mismo gráfico. Esta representación se basa en el producto escalar entre vectores fila y vectores columna, por lo tanto depende más de las longitudes y de los ángulos formados por los vectores que de las distancias entre puntos. Dada una matriz rectangular de datos, el biplot representa a la misma en pocas dimensiones, donde los vectores observaciones $X_{1...i}$ y los vectores vértices $Y_{1...j}$ se interpretan específicamente como puntos, lo que se busca es que los valores de los productos escalares $X_i Y_j$ entre los correspondientes vectores fila y los vectores columna se aproximen tan exactamente como sea posible a los respectivos valores de la celda correspondiente de la matriz de contingencia. Un biplot modeliza los datos como la suma de un producto escalar en algún subespacio de baja dimensionalidad minimizando el error residual.

Las variables y observaciones son graficadas como vectores desde el origen o centroide. Las distancias entre puntos filas miden la discrepancia entre perfiles filas. Los puntos filas muy cercanos en el gráfico, tienen similar perfil fila. Distancias desde el origen (centroide) indican la discrepancia entre los perfiles filas y el centroide fila o la distribución fila marginal. Cuando las longitudes de los vectores son similares el gráfico sugiere contribuciones similares de cada variable en la representación realizada. El mismo tipo de interpretación puede realizarse sobre los perfiles columnas.

La distancia entre símbolos representando observaciones y símbolos representando variables no tiene interpretación, pero las direcciones de los símbolos desde el origen sí pueden ser interpretadas. Puntos filas y columnas que caen en la misma dirección respecto al origen se encuentran positivamente correlacionados, cuanto más cerca se halle un perfil fila de un vértice más se identifica este perfil con el grupo representado por el vértice, mientras que aquellos que caen en direcciones opuestas se encuentran negativamente correlacionados, pudiendo tener valores relativamente bajos. Las direcciones pueden cambiar si se grafican otras dimensiones, por lo que es importante realizar el análisis sobre un espacio con alta inercia.

Por otro lado, los ángulos entre los vectores que representan las variables pueden ser interpretados en términos de las correlaciones entre variables. Ángulos de 90° entre dos variables indican que ambas variables no se encuentran correlacionadas. Alejamientos de este

valor (tanto sea en valores menores como mayores a 90°) implican correlación (positiva o negativa). Es decir un ángulo cercano a cero implica que ambas variables están fuertemente correlacionadas en forma positiva y un ángulo cercano al ángulo llano entre dos variables indica que ambas muestran fuerte correlación negativa.

En la figura 4 presentamos un biplot correspondiente al caso en tratamiento. En la misma, se representaron los puntos correspondientes a las filas (color rojo) y a las columnas (color azul) y los ejes principales. Es habitual representar el primer eje principal horizontalmente (eje X) y el segundo eje principal verticalmente (eje Y), que según vimos en la tabla 8 explican el 96,45% de inercia.

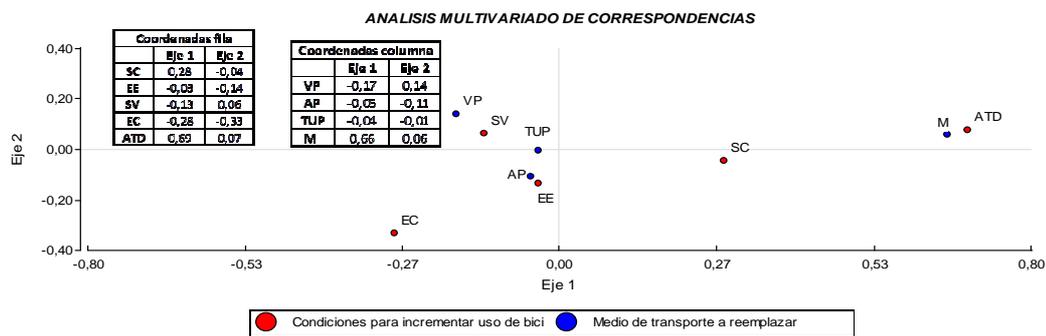


Figura 4. Análisis de correspondencia, relacionando variables, medios de transporte a reemplazar y observaciones, condiciones para incrementar el uso de bicicleta.

Del análisis de la figura 4 se observa que:

Considerando las distancias al centroide y entre sí de los medios, TUP y AP, son los que más aportarían al medio bicicleta en porcentajes similares, luego VP y prácticamente M no aportarían usuarios. Considerando las distancias al centroide y entre sí de las mejoras, EE y SV son las que más influyen para la posible migración al medio bicicleta, luego SC y prácticamente no influyen EC y ATD.

Si proyectamos sobre el eje principal horizontal y analizamos mejoras vs medios se puede observar que SV es la que captaría porcentajes similares de TUP, AP y VP; EE, de TUP y AP, en menor medida de VP.

Si eliminamos la modalidad M, el espacio que nos queda definido tendrá dos dimensiones (número de columnas-1). En la tabla 9 se observa que las proyecciones de los cinco perfiles fila explican el 100,00% de la inercia total. Es conveniente eliminar aquellas variables que no sean representativas para tener una mejor comprensión de la representación gráfica.

Tabla 9. Contribución a la chi-cuadrado por eje sin usuarios de Moto.

Contribución a la Chi cuadrado					
	Autovalor	Inercias	Chi-Cuadrado	(%)	% acumulado
1	0,11	0,010	2,73	75,7	75,7
2	0,06	0,004	0,88	24,3	100

4. CONCLUSIONES Y RECOMENDACIONES

Se realizaron encuestas para determinar variables (mejoras), que modificadas, permitan canalizar los desplazamientos hacia vehículos no motorizados, bicicleta, como medio de transporte alternativo desde otros medios.

Para la interpretación de los resultados de las encuestas se aplicó un análisis multivariado de correspondencias. El análisis de correspondencias (AC) constituye una herramienta de principal importancia para el análisis de variables categóricas con los que se construyen tablas de contingencia. Opera sobre la matriz de desviaciones Chi cuadrado generando un subespacio óptimo donde se pueden representar gráficamente las filas (observaciones) y columnas (variables) de la tabla de contingencia, ponderadas por sus respectivos pesos, como puntos en un espacio de baja dimensión (generalmente bidimensional) para así poder identificar las asociaciones de mayor peso entre las modalidades de varias variables cualitativas.

La representación gráfica utilizada fue el biplot. En él se muestran todas las observaciones y variables como puntos graficados un solo plano, de forma tal que se pueden hacer interpretaciones sobre las relaciones conjuntas. Las variables son graficadas como vectores desde el origen. En esta representación se interpreta las distancias entre puntos fila y la distancia entre el punto fila y el centroide. Las distancias entre puntos fila y columnas carece de sentido, pero aquellos que caen en la misma dirección se encuentran positivamente correlacionados, mientras que aquellos que caen en direcciones opuestas se encuentran negativamente correlacionados.

A modo de ejemplo se consideraron sólo dos variables, medios de transporte que los usuarios tendrían preferencia a reemplazar por bicicletas (cuatro modalidades) y que mejoras habría que implementar en el medio bicicleta para que se diera tal migración (cinco modalidades). Se concluyó que un alto porcentaje (89,48%) de los usuarios de los medios TUP, VP y AP tendrían preferencia a migrar al modo bicicleta si se mejoraran las condiciones de SV, EE, SC, EC y ATD. Sólo el 10,52% de usuarios del medio M migrarían si se dieran esas mejoras. Esto se ve reflejado en la figura 4.

Como desventaja, dado que para graficar la tabla de contingencia se reduce la dimensionalidad, se pierde precisión en la interpretación de las relaciones conjuntas, para una mejor comprensión se debe acompañar el biplot con la ejecución del modelo de tabla 6.

No se encuentran antecedentes de la aplicación de esta técnica a encuestas sobre transporte urbano.

REFERENCIAS BIBLIOGRAFICAS

- Greenacre, M. (2008) *La práctica del análisis de correspondencias*. Editora, Fundación BBVA (2008).
- Di Rienzo J.A., Casanoves F., Balzarini M.G., Gonzalez L., Tablada M., Robledo C.W.(2008). *InfoStat, versión 2008*, Grupo InfoStat, FCA, Universidad Nacional de Córdoba, Argentina.
- G.I.U.N.C. (2012,2013) *Modos de transporte no motorizado en la ciudad de Córdoba. Caso bicicleta*. Informes parciales Grupo de investigación de la Maestría en Ciencias de la Ingeniería-Mención en Transporte. FCEfYN. SECyT.Universidad Nacional de Córdoba. Argentina.
- Pearmain D., Swanson J., Kroes E., Bradley M. (1991) *Stated preference Techniques: a guide to practice*. STEER DAVIES GLEAVE, HAGUE CONSULTING GROUP.
- PTUMA (2009) *Encuesta de Origen / Destino, Movilidad en el Área Metropolitana de Córdoba*. Proyecto de Transporte Urbano en Areas Metropolitanas. <http://www.ptuma.gob.ar/publicaciones/index.html>.