



---

# **Estado de la Práctica en Encuestas de Movilidad Metropolitanas y su Importancia en Planificación Estratégica**

Juan de Dios Ortúzar

Departamento de Ingeniería de Transporte y Logística  
Centro de Desarrollo Urbano Sustentable (CEDEUS)  
Pontificia Universidad Católica de Chile



## Encuestas Origen-Destino Metropolitanas

---

- El estado de la práctica en cuanto a recolección de datos para estimar y mantener un modelo de planificación estratégica de transporte y uso de suelo, considera:
  - Diseñar un proceso de recolección de información continua; esto es, considerando cada día de la semana, durante todo el año y por varios años
  - Recolectar información sobre *todas* las etapas de cada viaje, tal que sea posible relacionar modo, localización, hora del día y longitud de los viajes
  - Considerar *todos* los modos, incluyendo viajes no motorizados, y recolectar información sobre *todos* los miembros del hogar
  - Diseñar un sistema integrado; esto es, encuestas en hogar más entrevistas en líneas pantalla y cordón
  - Obtener datos de gran calidad (bajo sesgo de respuesta), que sean lo suficientemente robustos como para ser utilizados a nivel desagregado
  - Geocodificar los datos para que puedan ser usados a cualquier nivel de agregación
  - Actualizar periódicamente matrices y modelos.



## Actualización Periódica de Matrices y Modelos

---

- En relación a lo último, se debería pensar en actualizar las matrices de viaje para toda el área cada 12-18 meses (ya que refleja el crecimiento diferencial en diferentes partes del área de estudio), dependiendo del tipo de ciudad estudiada
- Por otro lado, la puesta al día de modelos posiblemente afecte el tipo de datos a recolectar; así, parece necesario actualizar periódicamente:
  - Tasas de generación y atracción de viajes
  - Partición modal, incluyendo modos no motorizados (pues refleja posibles impactos de distintas políticas de transporte)
  - Nivel de tráfico en la red (permite identificar el crecimiento diferencial del flujo en las redes primaria, secundaria, de acceso, y en redes locales)
  - Tendencias en la tasa de motorización y formación de hogares en distintas municipalidades.
- Las prioridades de actualización dependen de las políticas de uso de suelo y transporte consideradas, la necesidad de monitorear el sistema, los requerimientos de modelación, las tasas de cambio esperadas, y los costos de recolección asociados (incluyendo el costo social de perturbar a los usuarios del sistema de transporte).



## Componentes de una Encuesta de Movilidad

---

- Entrevistas en hogares (y lugares de trabajo)
- Encuestas de interceptación en cordones y líneas pantalla
  - Entrevistas a la vera del camino
  - Encuestas a bordo de vehículos
  - Encuestas en estaciones de intercambio.
- Aforos de tráfico, bicicletas y peatones
- Otros estudios relacionados:
  - Datos de uso de suelo, incluyendo distribución de empleo
  - Grupos focales, preferencias declaradas, análisis de actitudes
  - Encuestas a transportistas y embarcadores
  - Información sobre la red y sistema de transporte - incluyendo datos sobre rutas, tarifas, frecuencias y nivel de servicio
  - Estudios de tiempos de viaje.



## Viajes Interno-Interno e Interno-Externo

---

### ■ Encuestas en el Hogar

- Tamaño de muestra (tema de gran interés que veremos a continuación)
- Alto costo por viaje, a cambio de ser:
  - ▶ Muy útiles para modelar generación de viajes y tasas de motorización
  - ▶ Útiles para estimar modelos desagregados de elección modal y modelos de elección horaria
  - ▶ Prácticamente la única fuente de información para modelar viajes encadenados
  - ▶ De ayuda en la modelación de distribución de viajes, puesto que entregan información clave sobre longitudes de viaje.
- Rendimiento aproximado de un hogar por encuestador por día, a no ser que se aplique un sistema de auto-llenado importante
- Puede considerar envío por correo y sistemas mixtos (esto es en ocasiones obligatorio)
- No se debe olvidar a los no-residentes en hogares (hoteles, hospitales, colegios, cuarteles).



## Tamaño de Muestra para Encuestas en el Hogar

---

- Desafío del diseño: identificar estrategias y tamaños de muestra que permitan extraer conclusiones válidas, y estimar modelos confiables (insesgados) sin gastar una cantidad excesiva de recursos
- La esencia del cálculo del tamaño muestral es el compromiso:
  - Muestra muy grande → encuesta muy cara para los objetivos especificados y grado de precisión requerido
  - Muestra muy pequeña → resultados sujetos a tal variabilidad que decisiones no se pueden basar, con confianza, en los resultados del estudio.
- En el caso de estudios estratégicos, uso final de los datos será como entrada de diferentes modelos. Así, el cliente debiera ser capaz de especificar, para cada variable de interés:
  - Error estándar aceptable (cuánto podría variar la media estimada de ciertas variable clave, respecto a su media poblacional), y
  - Nivel de confianza aceptable (cuan a menudo la media estimada diferirá de la media poblacional).



## Tamaño de Muestra para Encuestas en el Hogar

---

- Por ejemplo, para modelar generación de viajes se requieren tamaños de muestra bastante bajos. Smith (1979) propone la fórmula:

$$n = \frac{CV^2 Z_\alpha^2}{E^2}$$

donde  $n$  es el tamaño de muestra,  $CV$  el coeficiente de variación ( $\sigma/\mu$ ) de la variable de interés,  $Z_\alpha$  el valor de la variable Normal estándar para el nivel de confianza  $(1-\alpha)$  requerido y  $E$  el nivel de exactitud expresado como proporción.

En el caso de la tasa de viaje promedio por hogar, experiencias anteriores sugieren que  $CV \sim 1,0$ ; además, se sabe que  $Z = 1,645$  para  $(1-\alpha) = 95\%$  y podemos utilizar  $E = 0,05$  (5% de precisión)

Reemplazando estos datos en la fórmula anterior se tiene:

$$n = 1.084$$

Esto es, una muestra de sólo 1.100 hogares permitiría estimar tasas de viaje con 5% de tolerancia el 95% de los casos.



## Tamaño de Muestra para Encuestas en el Hogar

---

- La situación cambia radicalmente si se desea estimar matrices O-D:
  - Smith (1979) muestra que se requeriría entrevistar a 4% de la población para estimar el número de viajes en celdas con más de 1.100 viajes al 90% de confianza y 25% de error
  - O sea, si hay menos de 1.100 viajes entre dos zonas, una muestra menor al 4% no sería suficiente para detectarlos !
- En Santiago, a nivel municipal (34 zonas), sólo 58% de las celdas O-D contenía más de mil viajes en 1991.
  - Así, asumiendo una buena tasa de respuesta (75%), como Santiago tenía 1.400.000 hogares en esa época → se necesitaría entrevistar 75.000 hogares
  - Y esto permitiría estimar matrices para 34 zonas con 25% de error al 90% de confianza
- ¿Es realmente necesario buscar un tamaño de muestra tan grande (y tan costoso) para alcanzar un objetivo tan pobre?





## Tamaño de Muestra para Encuestas en el Hogar

---

- Así, el tamaño de muestra parecería estar regido por la estimación de matrices a nivel zonal, pero:
  - Smith (1979) también demuestra que no es posible reducir el error a niveles aceptables en áreas con más de 100 zonas, ya que el tamaño de muestra requerido sería cercano al 100%
- Por tanto, si los objetivos del estudio incluyen la estimación de matrices O-D, se recomienda usar una combinación de encuestas en hogares e interceptación para aprovechar sus mayores eficiencias relativas en alcanzar distintos objetivos
- De esta forma, parece más conveniente diseñar el tamaño de la muestra en función de otros propósitos:
  - Estimar tasas de viaje por categoría de hogar, y/o
  - Tasas de motorización a nivel zonal, y/o
  - Elección modal por estrato de ingreso.
- Para todos estos propósitos, Santiago (con 6,5 millones de habitantes) requeriría aproximadamente 5 mil encuestas/año.



## Un Sistema Continuo de Recolección de Datos

---

Un sistema de recolección de datos continuo permite pensar en un equipo de terreno relativamente pequeño y muy bien entrenado, que logre alta calidad en los resultados; pero tiene algunos problemas a considerar:

- ¿Cómo mantener al personal en el tiempo y, al mismo tiempo, conseguir que sus niveles de motivación y rendimiento permanezcan al nivel deseado/requerido?
  - ¿Cómo lograr que las tasas de respuesta no bajen en el tiempo, debido a la saturación de los encuestados?
  - ¿Cómo se integran las muestras de años sucesivos cuando ocurre un cambio drástico (o masivo) en la oferta de transporte en la ciudad?
  - ¿Cuándo sería apropiado utilizar otros métodos (por ejemplo, internet) como parte de un proceso de recolección de información continua?
  - ¿Es necesario tener tamaños de muestra iguales cada año, o existen ventajas asociadas a tomar una muestra mayor cada tres o cuatro años y muestras menores en los años intermedios?
  - ¿Qué tipo de preguntas se necesitarán a futuro dado el esperado énfasis en asuntos como cambio climático?
- Finalmente, acá no estoy considerando un tema interesantísimo, ¿cómo integrar adecuadamente las nuevas fuentes de información (*big data*, *smart cities*) que seguramente revolucionaran la práctica?



## Corrección, Expansión y Validación

---

- Se debe garantizar que las distribuciones de tamaño de hogar y sexo/edad de la muestra, sean equivalentes a las de la población (típicamente datos censales)
- Esto requiere un enfoque iterativo. El método multi-proporcional converge en pocas iteraciones; además tiene la ventaja de no necesitar posteriormente el cálculo de factores de expansión
- Corrección por no respuesta de ítems:
  - En encuestas de auto-llenado se debe encuestar una muestra de validación usando entrevistas personales y ponderar los datos en forma adecuada
  - Con entrevistadores bien entrenados, este tipo de corrección no es requerida en el caso de encuestas con entrevista personal.
- En el caso de no respuesta completa de un hogar (o algún miembro), se debe diferenciar las situaciones de genuina no respuesta de la pérdida de muestra; en el primer caso, la corrección se basa en el número de visitas o recordatorios.



## Imputación

---

- La *no respuesta* dificulta identificar los parámetros poblacionales. De hecho, sólo es posible identificarlos haciendo ciertas hipótesis (normalmente no testeables) acerca de la distribución de los datos faltantes.
- No obstante, la *no-respuesta* no impide identificar rangos para los parámetros
- A veces es posible hacer inferencias – por ejemplo, una persona que no indica ser trabajador, pero reporta viajes al trabajo; pero esto no funciona para muchas variables
- Imputar implica substituir valores por datos faltantes en base a ciertas reglas o procedimientos; existen diversos métodos de imputación en la práctica y sólo los menciono como anécdota:
  - Imputación deductiva
  - Uso de clases promedio
  - Imputación “hot o cold-deck”, etc.

De hecho, es toda un área de investigación.



## Imputación y Expansión de la Información

---

- Finalmente, las variables imputadas pueden producir errores de estimación. Por esto, hoy se recomienda hacer *imputación múltiple*, combinando los estimadores de los modelos resultantes en cada caso para obtener valores que incluyen un error asociado.

## Expansión de los Datos

- Consiste simplemente en multiplicar las observaciones por el inverso de la tasa de muestreo. Pero en algunos casos este cómputo puede no ser trivial:
  - Por ejemplo, si el marco muestral está constituido por las direcciones del servicio de impuestos internos, es posible (aunque esto es más común en países en desarrollo) muestrear direcciones no elegibles (i.e. no residencial o demolida), o donde nadie conteste.



## Validación

---

- En estudios de movilidad contemporáneos, la validación más importante no depende de datos secundarios (ya que cada tipo de encuesta tiene sesgos particulares que confunden el proceso), sino que es bastante interna a cada encuesta
- Por validación normalmente se entiende:
  - Verificar la completitud y coherencia de los datos en terreno, y además codificarlos y digitalarlos en gabinete
  - Efectuar una validación computacional de rangos aceptables para las variables a fin de garantizar su coherencia interna; al finalizar ambos procesos se cuenta con “datos limpios” ... y
  - Usar información independiente para probar los datos expandidos:
    - ▶ Aforos en cordones y líneas pantalla
    - ▶ Viajes registrados en torniquetes del Metro
    - ▶ Número de alumnos en distintos colegios
    - ▶ Distribución del ingreso, etc.



## Datos Comparativos Ultimas Encuestas en Santiago

---

<b>Año</b>	<b>Hogares</b>	<b>Autos</b>	<b>Viajes por hogar</b>	<b>Viajes por persona</b>
<b>1977</b>	649.820	208.263	4,65	0,87
<b>1991</b>	1.162.845	418.624	4,98	1,29
<b>2001</b>	1.513.938	855.057	6,66	1,81
<b>2012</b>	2.051.300	1.160.700	5,55	1,81

Considerando sólo viajes motorizados

Notar que el 2001 en Santiago había, en promedio, 3,64 hab/hogar y que esta cifra bajó a 3,23 hab/hogar en 2012.



## Datos Comparativos Ultimas Encuestas en Santiago

---

<b>Año</b>	<b>Transporte Privado (%)</b>	<b>Transporte Público (%)</b>	<b>Otros (%)</b>
<b>1977</b>	11,60	83,36	5,04
<b>1991</b>	19,69	70,54	9,77
<b>2001</b>	38,96	51,76	9,28
<b>2012</b>	46,4	46,9	6,70

Considerando sólo viajes motorizados





## Datos Comparativos Ultimas Encuestas en Santiago

---

<b>Partición modal (%)</b>	<b>2001</b>	<b>2012</b>
<b>Auto</b>	21,0	26,1
<b>Bip</b>	30,1	23,7
<b>Taxi Colectivo</b>	2,5	2,4
<b>Taxi</b>	1,2	1,8
<b>Otros</b>	3,6	5,3
<b>Caminata</b>	38,3	34,6
<b>Bicicleta</b>	2,1	3,9



## Datos Comparativos Ultimas Encuestas en Santiago

---

En la EOD-2001 se gastaron aproximadamente 500 mil horas hombre en total:

- 1 gerente, 2 codificadores, 2 revisores/validadores y 25 encuestadores en cada uno de los 4 centros de operación.

Con esto, el costo total del estudio fue  $\approx$  US\$ 1,2 millones:

- encuesta en hogares: 2,10 US\$/viaje;
- encuesta de interceptación: 0,36 US\$/viaje

No se ha hecho, que yo sepa, este tipo de cálculo para la Encuesta 2012, pero se sabe que costó aproximadamente US\$ 1,26 (695 millones de pesos).



## Datos Comparativos Ultimas Encuestas en Santiago

	Período Normal	Período Estival	Total 2001	Total 2012**
Hogar completo	12.346	3.191	15.537	18.320
Hogar incompleto	338	3	341	
Rechazos	2.653	633	3.286	8.256
Falso (descartado)	101	-	101	
Se recibió cuestionario pero no se devolvió	0	4	4	
Cuestionario devuelto vacío	0	1	1	
No hubo contacto*	2.679	924	3.603	746
Tasa de respuesta (%)	68,15	67,09	67,93	67,0

\* Notar que como muchos de estos pueden ser no-elegibles, la tasa de respuesta real debe ser más alta

\*\* Esto corresponde a los datos “casi finales” por lo que podría haber variaciones



## Datos Comparativos Ultimas Encuestas en Santiago

	Total 2001	Total 2012	Razón del rechazo	
Hogar completo	15.537	18.320	No tiene tiempo	4.182
Hogar incompleto	341		No cree en confidencialidad	607
<b>Rechazos</b>	<b>3.286</b>	<b>8.256</b>	Preguntas muy personales	479
Falso (descartado)	101		Encuestas no sirven	518
Se recibió cuestionario y no se devolvió	4		No gana nada al responder	445
Cuestionario devuelto vacío	1		Le aburre	358
No hubo contacto*	3.603	746	No da razón	1.667
Tasa de respuesta (%)	67,93	67,0		