**MODELOS DE GENERACIÓN DE VIAJES**

*Segundo informe sobre una muestra de datos provenientes de la Encuesta de Origen y Destino de Santiago de Chile del año 2012*

Autor: José Reyes Saldías

Fecha de 1º envío: 6 de mayo de 2021

Fecha 2º envío: 16 de mayo de 2021

**Tarea 2 – Modelos de Generación de Viajes**

Curso: Modelos de Demanda de Transporte

**PAUTA INFORME TAREA 2**

|  |  |  |  |
| --- | --- | --- | --- |
| **ASPECTOS FORMALES (10%)** | **Máx** | **% de total** | **Puntaje** |
| Redacción y ortografía | 2,0 | 3,3 |  |
| Cumplimiento del formato (cumple o no) | 2,0 | 3,3 |  |
| Presentación gráficos y figuras | 2,0 | 3,3 |  |
| **TOTAL** | **6,0** | **10%** |  |
| **ANÁLISIS ESTADÍSTICO ( %)** | **Máx** | **% de total** | **Puntaje** |
| Análisis de inconsistencias (1,2)  Análisis estadístico y/o gráfico de variables (2,4)  Categorización (2,4) | 6,0 | 15,0 |  |
| **TOTAL** | **6,0** | **15%** |  |
| **MODELACIÓN ( %)** | **Máx** | **% de total** | **Puntaje** |
| Explicación de la obtención del mejor modelo – Propósito de trabajo (0,75)  Estimación adecuada de modelos de RLM – Propósito de trabajo (0,75) | 1,5 | 15,0 |  |
| Explicación de la obtención del mejor modelo – Propósito de estudio (0,75)  Estimación adecuada de modelos de RLM – Propósito de estudio (0,75) | 1,5 | 15,0 |  |
| Explicación de la obtención del mejor modelo – Propósito de otro (0,75)  Estimación adecuada de modelos de RLM – Propósito de otro (0,75) | 1,5 | 15,0 |  |
| Correcto uso de tests (test-*t*, R2 , test-F) (1,0)  Consistencia de signos (0,5) | 1,5 | 15,0 |  |
| **TOTAL** | **6,0** | **60,0%** |  |
| **POLÍTICA PÚBLICA ( %)** | **Máx** | **% de total** | **Puntaje** |
| Predicción de política pública adecuado y coherente, utilizando variables adecuadas | 6,0 | 15,0 |  |
| **TOTAL** | **6,0** | **15,0%** |  |

**Nota:\_\_\_\_\_\_\_\_**

**MODELOS DE GENERACIÓN DE VIAJES**

*Segundo informe sobre una muestra de datos provenientes de la Encuesta de Origen y Destino de Santiago de Chile del año 2012*

**1. Preámbulo**

El presente informe tiene un doble objetivo, por un lado, busca modelar los factores que inciden en la generación de viajes basados en el hogar, considerando una agregación a nivel de hogares y basándose en una sub-muestra aleatoria de datos extraída de la Encuesta Origen Destino del año 2012 de Santiago. Por otro lado, en base a dicha modelación, se busca predecir cuál sería la incidencia que, sobre la generación de viajes, tendría la implementación de una política pública cuyo fin es promover la realización de estudios formales por parte de la población inactiva, sean estos jubilados y/o desempleados. Tanto en la modelación como la predicción de viajes se distinguen entre tres propósitos: viajes al trabajo, viajes al estudio y otros viajes.

Lo que sigue a continuación se organiza en cuatro secciones: la primera de ellas describe la base de datos, junto con las correcciones y la fundamentación sobre la que se articula la presente modelación de viajes. La segunda sección describe el proceso con el cual se modeló la generación de viajes, además de presentar los principales resultados de ésta y evaluar la calidad del ajuste de los modelos obtenidos. La tercera sección muestra los resultados que, a partir de la implementación de la política pública indicada, se esperarían sobre los patrones de generación de viajes. Por último, el cuarto apartado cierra con algunas reflexiones sobre el estudio.

Finalmente, debe hacerse mención que este documento constituye el segundo reporte de una serie de cuatro publicaciones referidas a los patrones de movilidad en Santiago y a las técnicas para su modelación.[[1]](#footnote-1)

**2. Descripción de los datos y fundamentos de la modelación**

**2.1 Datos y correcciones**

Los datos sobre la generación de viajes con que se cuenta en este análisis provienen de la encuesta de Origen y Destino efectuada el año 2012 en Santiago. Esta es una encuesta que se realiza en hogares durante un día en específico de la semana y recoge la información sobre los viajes en la vía pública que realizan todos los miembros del hogar durante dicho día. Para este estudio se seleccionó una sub-muestra aleatoria de 6.000 hogares. En base a dicha sub-muestra se cuenta con datos a nivel del hogar del total de viajes generados y según el tipo de propósito, a la vez que se cuenta con información sociodemográfica de los hogares. También se posee información relativa a la propiedad de bienes y/o facilidades para la movilidad, tales como cantidad de autos, bicicletas, licencias de conducir y pases escolares en el hogar.

Para efectos de corregir la base de datos, se verificó que los totales de personas por hogar correspondieran con la suma de la cantidad de personas dentro de cada tramo de edad. Asimismo, se verificó que el total por género en el hogar sumara la cantidad total de personas en el hogar. También, se verificó que la cantidad de licencias de conducir no fuese superior a la cantidad de adultos viviendo en el hogar. Asimismo, se constató la presencia de posibles datos fuera de los rangos habituales, en donde se consideró como punto de corte una cantidad de viajes de trabajo y de estudios superiores a cinco veces la cantidad de trabajadores y estudiantes en el hogar, respectivamente. En una gran medida, las observaciones cumplieron con los criterios señalados. No obstante, fueron sacadas de la muestra, además de los casos ya indicados, observaciones en las cuales el número total de bicicletas en el hogar no estaba especificado. Así, de 5875 observaciones, la muestra se redujo a 5.875, una diferencia del orden del 2%.

**2.2 Categorización de los datos**

Tal como se mencionó, la base de datos cuenta con información sobre el ingreso del hogar, el número de vehículos y la cantidad de bicicletas, tanto de niños como de adultos. Asimismo, se cuenta con el número de pases escolares y de licencias de conducir en el hogar. Todas estas variables se espera que tengan una incidencia positiva sobre la cantidad total de viajes generados. Ahora bien, al distinguir entre el tipo de viaje, se esperaría que el número de vehículos en el hogar y la cantidad de licencias de conducir tengan una mayor incidencia sobre los viajes al trabajo y a viajes con otros propósitos, y no tanto sobre los viajes con motivo de estudios. A su vez, se esperaría que dicha incidencia aumentase a una tasa decreciente. Esto se debe que, tal como se aprecia en la **Figura 1** y en la **Figura 2**, la distribución del ingreso y de la cantidad de vehículos se concentra en el margen cercano a cero, por lo que posiblemente la incidencia de esta variable se reduzca proporcionalmente al ir aumentando su valor.

**Figura 1**: Tabla de frecuencias de cantidad de Vehículos. Fuente: elaboración propia.



**Figura 2**: Histograma de los Ingresos por Hogar. Fuente: elaboración propia.



De manera opuesta, la cantidad de pases escolares incidiría más fuertemente sobre los viajes con motivo de estudios y a otros viajes, y no tanto sobre los viajes de trabajo. Lo mismo podría decirse acerca de la cantidad de bicicletas para niños. En cuanto a la cantidad de bicicletas para adultos, esta variable debería tener una incidencia más fuerte sobre la generación de viajes con otros propósitos. El ingreso, por último, debería estar impactando positivamente a los tres propósitos de viajes, aun cuando y como ya se mencionó, a una tasa decreciente.

En cuanto a la información demográfica y ocupacional del hogar, desafortunadamente los datos con que se trabaja presentan un nivel de agregación que dificultan seriamente la categorización de las observaciones. Así, por dar algunos ejemplos, se cuenta con el número de personas adultas y con el número de trabajadores por hogar, pero se desconoce cuántos de los adultos que hay en el hogar efectivamente trabajan y/o realizan otras actividades. De igual manera, y por dar otro ejemplo, no es posible identificar si los adultos de hasta 25 años en el hogar trabajan y/o estudian, ni tampoco discriminar directamente por género. Por lo tanto, los efectos cruzados que puedan obtenerse de las variables de interés deben interpretarse con extrema cautela.

En relación a la incidencia de estas variables, si bien las variables demográficas y ocupacionales con que se cuenta deberían tener un efecto positivo sobre la cantidad de viajes generados -ya que, tal como están medidas, al incrementarse las cantidades observadas en cualquiera de ellas, dejando a las demás constantes, no podría esperarse una disminución en los viajes generados- ello no necesariamente podría verificarse de manera clara en la modelación, particularmente debido a la altísima correlación que en principio presentarían estos datos debido al modo en cómo fueron agregados. Por lo tanto, no sería de extrañar que la cantidad de adultos mayores y de niños aparente tener una incidencia negativa sobre la cantidad de viajes generados, de igual manera que los jubilados, desempleados, discapacitados y dueños de casa, esto particularmente para los propósitos de trabajo y de estudios. En cambio, para otros propósitos de viajes esta incidencia podría presentarse positiva con una mayor posibilidad. Por último, la cantidad de mujeres en el hogar se esperaría que incida negativamente sobre los viajes de trabajo y positivamente sobre viajes con otros propósitos, aun cuando si no se encuentra esta relación no necesariamente habría que rechazar la existencia de ella.

Habiendo hecho las cautelas necesarias, para efectos de los datos con que se cuenta se trabajarán de manera logarítmica las siguientes variables: Ingreso, Número de vehículos, Número de bicicletas de niños Número de licencias de bicicletas de adultos, Licencias de conducir y Pases escolares.[[2]](#footnote-2) Las variables demográficas y ocupacionales se trabajarán de manera lineal. No obstante, para diferenciar el efecto cruzado de las variables se incluirán los siguientes grupos:

* Jóvenes – Estudiantes
* Adultos Jóvenes – Estudiantes
* Adultos – Estudiantes
* Adultos Mayores – Estudiantes
* Adultos Jóvenes – Trabajadores
* Adultos – Trabajadores
* Adultos Mayores – Trabajadores
* Mujeres – Dueñas de Casa

La inclusión de tales grupos cruzados dice relación, en primer lugar, con que, junto con modelar la generación de viajes, se buscar estimar con mayor precisión el impacto que tendría la política pública de promoción de estudios por parte de personas inactivas. En segundo, para evaluar la incidencia sobre la generación de viajes de grupos sociodemográficos claramente identificables, tales como los jóvenes-estudiantes, las mujeres-dueñas de casa y los/as adultos-trabajadores. Así, para cada uno de dichos grupos, además de los efectos lineales de las variables sobre los viajes generados, se incluirá como variable compuesta por la raíz cuadrada de los productos de las variables que los componen.

**3. Modelación y principales resultados**

**3.1 Modelos Estimados**

Tal como se mencionó anteriormente, para modelar la generación de viajes para cada propósito se articuló una asociación lineal entre los viajes generados y todas las variables indicadas previamente, sumando a eso las raíces de los productos cruzados. Dicha modelación se traduce en un modelo de regresión lineal de mínimos cuadrados generalizados, el cual es en muchos aspectos equivalente al método de mínimos cuadrados ordinarios, pero ajustando cada observación de acuerdo a un factor de expansión y corrección por tamaño del hogar y representación sociodemográfica (por comuna y nivel de ingreso, entre otras variables).

En este punto conviene, eso sí, hacer una cautela: las observaciones del número de viajes por propósito en cada hogar están concentradas en valores positivos cercanos a cero. Ello se traduce en que los errores de predicción no vayan a sean simétricos, sino que tenderían a aumentar en la medida en que los valores predichos y/o observados de la variable dependiente aumente. De ese modo, no es posible garantizar la normalidad ni la heterocedasticidad en la distribución de los errores. Si bien dicha falencia podría ser corregida especificando otro modelo de estimación, para efectos de este estudio se procederá con el método de mínimos cuadrados generalizados, tomando las debidas cautelas al momento de interpretar los resultados.

Habiendo sentado lo anterior, los modelos para cada uno de los tres propósitos fueron estimados. Las variables finales incluidas en los modelos fueron aceptadas en la medida en que la estimación les arrojase una significancia estadística superior al 90%.[[3]](#footnote-3) El resumen estadístico del modelo obtenido para la generación de viajes de Trabajo se muestra en el **Cuadro 1**.

**Cuadro 1**: Resultado de la estimación del Modelo de Generación de Viajes al Trabajo. Fuente: elaboración propia.



En términos generales, los signos de las variables se ajustan a lo esperado: número de vehículos, ingreso del hogar, cantidad de adultos y trabajadores inciden positivamente en la cantidad de viajes con propósito de trabajo. Es de notar que para este modelo se debió abandonar la hipótesis de la incidencia del logaritmo del Ingreso y, en cambio, se incorporó el Ingreso del Hogar de manera lineal debido a que ajustaba mejor a los datos. Por otro lado, el hecho de encontrar desempleados, estudiantes y discapacitados en el hogar tiene una incidencia en sentido negativo. Ciertamente, la cantidad de trabajadores es la variable que más incide en este caso. Con respecto al R2 ajustado, este es de 0,636, un valor razonablemente alto para este tipo de modelos. El valor p de Fisher es estadísticamente significativo (cercano a cero).

El resumen estadístico del modelo obtenido para la generación de viajes de Estudio se muestra en el **Cuadro 2**.

**Cuadro 2**: Resultado de la estimación del Modelo de Generación de Viajes al Estudio. Fuente: elaboración propia.



En este caso, de manera elocuente, la cantidad de estudiantes es lo que más determina la cantidad de viajes de estudio. Si bien los coeficientes de los signos parecen ser los correctos, llama la atención que la cantidad de pases escolares tenga un signo negativo. No sería de extrañar que ello se deba a la alta correlación entre las variables lo cual impida identificar con claridad el efecto d esta variable. El R2 ajustado de la regresión es alto, de un 0,689, mientras que la regresión como un todo es estadísticamente significativa.

Por último, el resumen estadístico del modelo obtenido para la generación de viajes con Otros propósitos se muestra en el **Cuadro 3**.

**Cuadro 3**: Resultado de la estimación del Modelo de Generación de Viajes con Otros propósitos. Fuente: elaboración propia.



A partir de los resultados del **Cuadro 3**, se aprecia que dueños/as de casa, desempleados y jubilados generan más viajes de este tipo. Los estudiantes también parecen incidir positivamente sobre este tipo de viajes. El R2 ajustado, como suele ser habitual para este tipo de viajes es muy bajo, por lo que el modelo es poco lo que logra explicarlos, aun cuando la regresión como un todo sea significativa.

**3.2 Modelo de Predicción**

Como una manera de evaluar la capacidad predictiva de los modelos, a continuación, se presentan los resultados de estos al ser aplicados en un nuevo conjunto de 6.000 observaciones. Es de notar que tales observaciones también provienen de la Encuesta Origen Destino de Santiago del año 2012 y, además, fueron corregidas a través del mismo proceso que los datos con que fue calibrado el modelo. De ese modo, la muestra se redujo a un total de 5.873 observaciones. Los resultados de la predicción para los tres modelos se muestran en el **Cuadro 4**.

**Cuadro 4**: Resultado de la estimación del Modelo de Generación de Viajes con Otros propósitos. Fuente: elaboración propia.



Además, en la **Figura 3** se muestra el diagrama de dispersión entre los valores predichos y los valores observados de Viajes al Trabajo en la base de datos de predicción.

**Figura 3**: Diagrama de dispersión entre valores predicho y valores observados de Viajes al Trabajo en la base de datos de Predicción. Fuente: elaboración propia.



A su vez, la **Figura 4** muestra la distribución de los errores para los viajes con motivos de trabajo.

**Figura 4**: Histograma de los errores de predicción para los Viajes al Trabajo en la base de datos de Predicción. Fuente: elaboración propia.



Tal como se había adelantado, de las **Figuras 3** y **4** es posible observar que los errores de predicción no parecen estar distribuidos homocedasticamente ni de acuerdo a la distribución normal. Ello ratifica la cautela con que deben considerarse estos resultados en este estudio.

**4. Incidencia de la Política Pública sobre la generación de viajes**

**4.1 Descripción de la Política Pública**

Habiendo estimado los modelos para la generación de viajes para cada propósito, es posible evaluar la incidencia que tendría la implementación de una determinada política pública sobre ellos. Así, a continuación se describen los alcances de la política pública a evaluar para luego poder estimar correctamente el impacto de ella.

La política pública en cuestión se denomina “**Promoción de la capacitación y la enseñanza en adultos mayores, desempleados e inactivos**”. La política se fundamente en el hecho de que en Chile existe una proporción relativamente elevada de la población que se encuentra inactiva, sea porque no ha logrado encontrar un espacio dentro de las dinámicas laborales y/o sea porque sus competencias han quedado obsoletas debido a los cambios socioeconómicos de las últimas décadas. Así, tales personas pasan a convertirse en cargas para sus redes familiares y, en muchos, terminan sufriendo de la pérdida de su auto valoración, lo cual eventualmente los lleva a la depresión e incluso al suicidio. Dado eso, existe una Fundación que ha promovido la capacitación como medio para volver a insertar a las personas en las redes laborales, sean formales o no, para así mejorar su bienestar y calidad de vida, no sólo en lo económico sino también en lo emocional.

Puesto que el programa recién enunciado tiene unas posibilidades enormes de ser implementado, el actual constituye un momento ideal para evaluar *ex ante* el impacto que ésta podría tener sobre los patrones de viaje. La presente estimación se realiza a partir de los modelos calibrados anteriormente y considerando las variables presentes en la base de datos de predicción. Comenzando con el modelo para la generación de viajes al trabajo, el cual se resumió en el **Cuadro 1**, es posible identificar que las siguientes variables se verían afectadas: en el corto plazo, la cantidad de estudiantes y desempleados, mientras que en el largo plazo, la cantidad de trabajadores y la cantidad de adultos trabajadores. Del mismo modo, es posible estimar el efecto total para cada uno de los demás modelos, los cuales tal como se muestra en el **Cuadro 5**. Como efecto de base, se supuso que la política pública tendría una penetración en un 5% de los desempleados y de los jubilados. Ese porcentaje de la población pasaría a calificarse como estudiantes (aun cuando los jubilados no perderían su condición de jubilados). Asimismo, se considera que en el largo plazo, ese porcentaje pasaría a ser considerado como trabajados y adultos-trabajadores (tal como se definió en la especificación de los modelos).

**Cuadro 5**: Estimación del impacto de la política en la generación de viajes por tipo de propósito. Fuente: elaboración propia.



Como se aprecio en el **Cuadro 5**, la política generaría tanto aumentos como disminuciones en los viajes para cada propósito, variando también en el corto y en el largo plazo. Es importante, además, considerar que los efectos sólo consideran la duración de este programa durante un tiempo acotado. En caso de extenderse sus efectos, particularmente sobre los viajes de estudios, podrían persistir sus efectos incluso en el largo plazo.

**5. Propagación de Errores y palabras al cierre**

Una manera de verificar la robustez de los resultados obtenidos consiste en estimar la magnitud del error de predicción del modelo al suponer un determinado error al momento de medir las variables explicativas. Así, en este apartado se trabaja bajo el supuesto de que las variables explicativas poseerían un error de medición del 8%. Tomando como caso de referencial un hogar cuyas observaciones tomen el valor promedio de la muestra de la base de datos de calibración del modelo, los errores de medición se propagarían en las respectivas predicciones en las magnitudes que se muestran en el **Cuadro 6.**

**Cuadro 6**: Errores de predicción en base a errores de medición del 8%. Fuente: elaboración propia.



Así, se aprecia que los errores en la medición de las variables generarían que la predicción tuviese un margen de error del orden del 8% para los Viajes al Trabajo y al Estudio, mientras que los Viajes con Otro Propósitos tendrían un error del 3% aproximadamente. Posiblemente, la baja en dicho valor para Viajes con Otros Propósitos se deba al bajo valor predictivo que en general tiene dicho modelo (un R2 de 0,2).

Además de la propagación del error, es posible identificar cuáles variables sería conveniente medir con mayor precisión. Para ello, se muestra el **Cuadro 7**, el cual especifica cuál sería la disminución en el error de predicción en cada modelo al mejorar marginalmente la medición en cada variable. Se incluye, además, una columna con promedio las mejoras en los tres modelos ponderando de acuerdo al R2 de cada uno.

**Cuadro 7.** Mejora en la predicción en base a errores de medición del 8%. Fuente: elaboración propia.



Desafortunadamente, el **Cuadro 7** entrega valores negativos para algunas variables, por lo cual es altamente posible que éste contenga errores en su cálculo, por lo que tales valores deben ser tomados con cautela. De cualquier modo, de la lectura de dicho cuadro se deriva que la variable que más relevante sería medir correctamente es la cantidad de Estudiantes en el Hogar.

Por último, es importante resaltar que, si bien el análisis de la propagación de los posibles errores de medición es relevante, de igual importancia es la evaluación de la calidad de la especificación del modelo. Tal como ha sido indicado a lo largo de este reporte, el modo con que se agregaron los datos e, incluso, el mismo modelo utilizado generaron inquietudes con respecto a su idoneidad. En fin, se espera que en un próximo reporte estas dudas puedas ser subsanadas.

1. Para mayor información sobre el estudio, contactar al autor del documento, José Reyes, jpreyess@uc.cl. [↑](#footnote-ref-1)
2. En estricto rigor, se calcula el logaritmo de tales variables más 1, con el fin de no estimar logaritmos de cero. [↑](#footnote-ref-2)
3. Habitualmente se exige una significancia mínima del 95%, pero debido a la incertidumbre sobre la distribución de los residuos, en este estudio se relaja el nivel de aceptación para las variables. [↑](#footnote-ref-3)