

MODELO

Nuestro modelo tiene como finalidad explicar el Gasto en Transporte Público (mensual) generado por las familias de Santiago, basándonos en modelos teóricos realizados con anterioridad en estudios de la demanda por transporte público. Los datos son los recopilados en la encuesta Origen Destino de 1991 en una muestra de 29182 familias, realizada por la Comisión de Planificación de Inversiones en Infraestructura y de Transporte; Secretaría Ejecutiva (SECTRA), organismo gubernamental dedicado al estudio de estas materias.

Entonces Planteamos el Modelo:

$$GTP = - \text{INGRESO} - \text{AUTO} + \text{PERSONAS} + i$$

El GASTO (GTP) es una estimación del precio promedio en transporte público (Microbuses, Metro, Taxis colectivos urbanos) en valores de diciembre del año 2000 por la cantidad de viajes mensuales realizados por las familias encuestadas. La variable INGRESO es la marca de clase de 8 intervalos establecidos entre los cuales se reparte la muestra. Las PERSONAS corresponden al número de integrantes por familia. La variable AUTOS se refiere al número de autos en la familia.

El parámetro Intercepto del modelo.

El parámetro Corresponde al cambio en el gasto por cada peso extra de ingreso.

El parámetro Explica el cambio en el gasto a medida que aumenta la cantidad de autos.

El parámetro Explica el cambio en el gasto a medida que aumenta la cantidad de integrantes de la familia.

El parámetro i Residuos de la muestra (término de perturbación).

Entonces, al correr la regresión con estas variables y parámetros, nos queda:

LS // Dependent Variable is GTP

Sample: 1 29182

Included observations: 29182

Variable Coefficient Std. Error t-Statistic Prob.

C 210.4129 12.53868 16.78110 Prob.

INGRESO -7.45E-05 1.65E-05 -4.521931 0.0000

PERSONAS 222.5550 2.778711 80.09290 0.0000

AUTOS -210.1617 7.937419 -26.47734 0.0000

R-squared 0.198156 Mean dependent var 996.9543

Adjusted R-squared 0.198074 S.D. dependent var 887.7369

S.E. of regression 794.9715 Akaike info criterion 13.35675

Sum squared resid 1.84E+10 Schwarz criterion 13.35788

Log likelihood -236291.8 F-statistic 2403.541

Durbin-Watson stat 1.825131 Prob(F-statistic) 0.000000

De la regresión podemos ver que:

- Los signos de los coeficientes son coherentes con lo planteado
- Se rechaza la hipótesis nula de significancia individual de los parámetros, ya que la probabilidad de que estos sean cero es nula (Prob. 0.0000).
- La significancia global también es aceptada (F-statistic 2403.541)

Pero antes de continuar, para un estudio más significativo, dado que nuestro R² no es muy alto, es decir, las variables no explican poderosamente el gasto en transporte, desarrollaremos un nuevo modelo incluyendo dos nuevas variables:

$$GTP = - \text{INGRESO} - \text{AUTO} + \text{PERSONAS} + \text{EMPLEADOS} + \text{ESTUDIANTES} + i$$

Las nuevas variables corresponden a la cantidad de personas empleadas (EMPLEADOS) y estudiantes (ESTUDIANTES) en la familia.

El parámetro β_4 Corresponde al cambio en el gasto por cada persona empleada.

El parámetro β_5 Corresponde al cambio en el gasto por cada persona estudiante.

El parámetro ϵ_i Residuos de la muestra (término de perturbación).

La regresión nos queda:

LS // Dependent Variable is GTP

Sample: 1 29182

Included observations: 29182

Variable Coefficient Std. Error t-Statistic Prob.

C 264.0583 11.69404 22.58059 0.0000

INGRESO -0.000257 1.55E-05 -16.51458 0.0000

PERSONAS 73.39867 3.296594 22.26500 0.0000

AUTOS -253.8092 7.342384 -34.56768 0.0000

EMPLEADOS 309.7445 5.311703 58.31360 0.0000

ESTUDIANTES 307.0378 5.693779 53.92514 0.0000

R-squared 0.318796 Mean dependent var 996.9543

Adjusted R-squared 0.318679 S.D. dependent var 887.7369

S.E. of regression 732.7574 Akaike info criterion 13.19383

Sum squared resid 1.57E+10 Schwarz criterion 13.19554

Log likelihood -233912.7 F-statistic 2730.806

Durbin-Watson stat 1.838096 Prob(F-statistic) 0.000000

Si planteamos las hipótesis de que las nuevas variables no son significantes para el modelo tenemos: $H_0 : \beta = 0$; $\beta = 0$ y $H_1 : \beta \neq 0$; $\beta \neq 0$ Luego probamos con un test de hipótesis lineales conjuntas:

$$F = \frac{R_C^2 - R_R^2}{1 - R_C^2} \cdot \frac{n - k}{r} \cdot F^{0.95}(r, n - k)$$
$$F = \frac{0.318796 - 0.198156}{1 - 0.318796} \cdot \frac{29182 - 6}{2} \cdot F^{0.95}(2,)$$
$$F_C = 2583.508 > F_T = 3$$

Como el F calculado es mayor que le F tabulado, se puede decir que existe evidencia suficiente al 5% de significancia para rechazar las hipótesis conjuntas que afirman que las variables no son significativas para el modelo ($H_0 : \beta = 0$; $\beta = 0$), por lo tanto estudiaremos el segundo modelo planteado:

$$GTP = - \text{INGRESO} - \text{AUTO} + \text{PERSONAS} + \text{EMPLEADOS} + \text{ESTUDIANTES} + i$$

$$GTP = 264.0583 - 0.000257\text{INGRESO} - 253.8092\text{AUTOS} + 73.39867\text{PERSONAS} + 309.7445\text{EMPLEADOS} + 307.0378\text{ESTUDIANTES} + i$$

En la nueva regresión podemos ver que si bien se observa un R^2 bajo ("32%), esto no significa que el modelo no nos explique la variable dependiente (GTP); la significancia global que vemos con un valor igual a Prob(F-statistic) 0.000000 – nos indica, que aunque la fracción total de la varianza que puede explicar el modelo sea baja, la explica en forma significativa. Los signos son coherentes con lo propuesto, es decir, se cumplen las relaciones negativas del Ingreso y la cantidad de Autos en la familia. Las relaciones positivas de la cantidad de integrantes (Personas), Empleados y Estudiantes también son coherentes. Además encontramos que los parámetros en forma individual son significativos, ya que la probabilidad de rechazar la no-significancia de estos, es nula (Prob. 0.0000). La significancia global también queda demostrada por el F-estadístico alto, que al ser comparado con el F de tabla o tabulado entrega la evidencia suficiente para rechazar $\beta = 0$.

PROBLEMAS

Los problemas que podían surgir en la estimación del modelo eran entre otros: *autocorrelación*, *multicolinealidad* y *heteroscedasticidad*. Analizaremos estos dos últimos con respecto a nuestro modelo.

La multicolinealidad en nuestro modelo no esta presente ya que no muestra parámetros poco significativos

con un alto R², es decir, no existe dependencia entre las variables explicativas. La heteroscedasticidad (cambio en la variabilidad de los errores) la podemos probar a través de un test de White:

White Heteroskedasticity Test:

F-statistic 196.9372 Probability 0.000000

Obs*R-squared 1845.521 Probability 0.000000

Test Equation:

LS // Dependent Variable is RESID^2

Sample: 1 29182

Included observations: 29182

Variable Coefficient Std. Error t-Statistic Prob.

C 31011.78 28758.76 1.078342 0.2809

INGRESO 0.329092 0.057941 5.679764 0.0000

INGRESO^2 -2.16E-07 3.35E-08 -6.456703 0.0000

PERSONAS 93075.10 12956.05 7.183909 0.0000

PERSONAS^2 462.6478 1257.799 0.367823 0.7130

AUTOS -17716.58 18628.68 -0.951038 0.3416

AUTOS^2 7579.446 6797.370 1.115056 0.2648

EMPLEADOS -54787.86 16929.50 -3.236236 0.0012

EMPLEADOS^2 30284.88 4086.904 7.410225 0.0000

ESTUDIANTES 28435.84 18136.34 1.567893 0.1169

ESTUDIANTES^2 242758.21 6674.740 6.405974 0.0000

Bajo las hipótesis H₀ : Homocedasticidad vs H₁ : Heterocedasticidad, el test de White nos muestra que un – F-statistic 196.9372 – indica imposible aceptar H₀, por lo tanto el modelo presenta heteroscedasticidad.

Una medida de corrección frente a la presencia de heteroscedasticidad es tratar de mejorar la especificación usando variables dicótomas entre los cortes transversales, por esto plantemos el siguiente modelo:

GTP = - INGRESO - D1AUTO + PERSONAS + EMPLEADOS + ESTUDIANTES

Donde D1 vale 1 cuando la familia posee 1 o más autos y 0 cuando la familia no tiene autos. La regresión de este nuevo modelo con el test de White es:

White Heteroskedasticity Test:

F-statistic 225.0821 Probability 0.000000

Obs*R-squared 1894.853 Probability 0.000000

Test Equation:

LS // Dependent Variable is RESID^2

Sample: 1 29182

Included observations: 29182

Variable Coefficient Std. Error t-Statistic Prob.

C 27503.97 28864.93 0.952851 0.3407

INGRESO 0.351649 0.058050 6.057724 0.0000

INGRESO^2 -2.06E-07 3.34E-08 -6.158900 0.0000

PERSONAS 95494.10 13011.53 7.339191 0.0000

PERSONAS^2 89.38743 1263.517 0.070745 0.9436

DAUTOS -3389.383 14809.04 -0.228873 0.8190

EMPLEADOS -67058.36 17001.16 -3.944340 0.0001

EMPLEADOS^2 35340.18 4097.503 8.624807 0.0000

ESTUDIANTES 26328.95 18218.53 1.445174 0.1484

ESTUDIANTES^2 44616.46 6706.294 6.652923 0.0000

Como lo muestra el test de White hecho a la nueva regresión, la heteroscedasticidad no fue corregida con el uso de variables dicótomas (ver anexo n°2) ya que el valor del F-statistic 225.0821 sigue siendo tan alto como el anterior, siendo imposible aceptar nuevamente la hipótesis de homocedasticidad. Frente a esto, otra solución posible al problema de heteroscedasticidad, es el uso del Método de Mínimos Cuadrados Ponderados, pero que dejaremos como la solución propuesta por estar fuera del alcance del curso.

En conclusión,

BIBLIOGRAFIA

Modelling Transport; Juan de Dios Ortúzar, 1994

Modelos Econométricos de Elección Discreta; Juan de Dios Ortúzar, Ediciones PUC, 2000

Modelos de Demanda de Transporte; Juan de Dios Ortúzar, Ediciones PUC, 1994

Comisión de Planificación de Inversiones en Infraestructura y de Transporte; Secretaría Ejecutiva (SECTRA)

Modelos de Demanda de Transporte; Juan de Dios Ortúzar, 1994

No tomaremos autocorrelación como un problema significativo en nuestro modelo, ya que esta se encuentra presente principalmente en análisis de series de tiempo y no en datos de corte transversal.