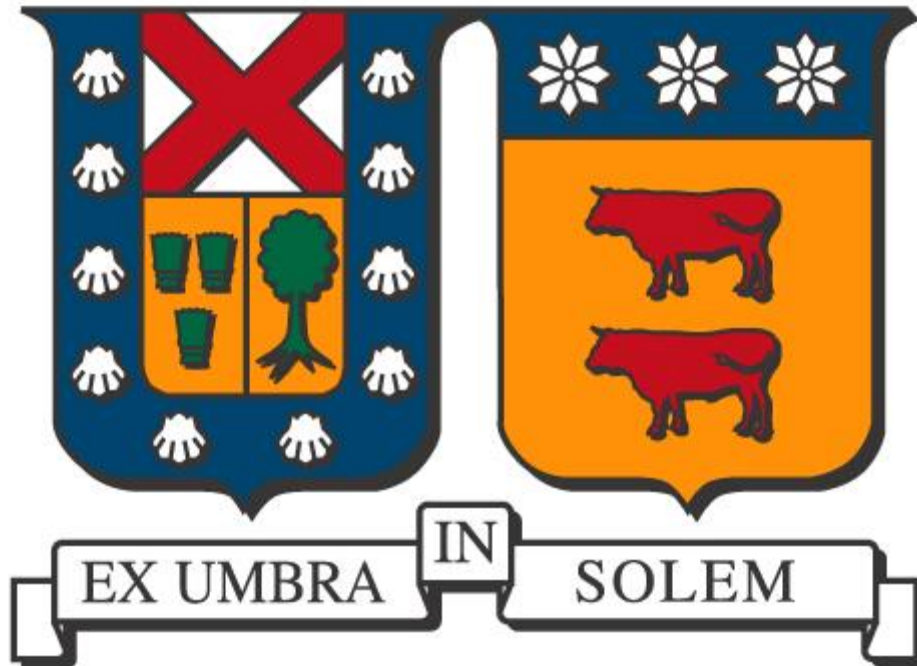


UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA

DEPARTAMENTO DE INGENIERÍA COMERCIAL

SANTIAGO - CHILE



**ESTUDIO ECONOMETRICO SOBRE FACTORES QUE EXPLICAN EL NIVEL DE  
INGLÉS DE LOS CHILENOS**

**SERGIO PATRICIO DOMEQC MUSIATE**

MEMORIA PARA OPTAR AL TÍTULO DE  
INGENIERO COMERCIAL

PROFESOR GUÍA : SR. RODRIGO ORTEGA

PROFESOR CORREFERENTE: SR. LUIS ACOSTA

MARZO 2022

## **Resumen**

El presente trabajo consiste responder la pregunta: ¿qué factores influyen en el nivel de inglés de los chilenos?. Esto mediante al análisis mediante regresión lineal a los datos obtenidos por una encuesta aplicada a 125 personas (111 respuestas que sirvieron, luego de aplicadas técnicas estadísticas bajó a 85, lo que nos da un 10,65% de error y 95% de significancia). Las principales variables que influyen de manera positiva en el nivel de inglés (aproximado por puntaje de los tests TOEIC, TOEFL, IELTS o Pearson Academic PTE) son género (hombres tienen mayor nivel), el haber estudiado en colegio particular subvencionado, haber llegado hasta enseñanza media completa, que su madre haya llegado hasta enseñanza media completa o universitaria completa y residir en Peñalolén cuando se aprende inglés. Sorprendentemente, la edad, el haber tenido al menos una asignatura dictada 100% en inglés, el ingreso por persona en el hogar, el haber vivido en países anglosajones, el haber estudiado para el examen correspondiente, el ser autodidacta y tener personas con las que conversar en inglés no afectan el nivel de éste.

## **Abstract**

This research consists in answering the question: which factors influence the level of english of chileans? through a linear regression analysis to data obtained via a survey applied to 125 people (111 useful answers and after statistical techniques applied, it went down to 85, which gives a 95% of significance and 10.65% of error). The principal variables that influence in a positive way the level of english of chileans (approximated by the score of TOEIC, TOEFL, IELTS and Pearson Academic PTE tests) are gender (men perform better), to have studied in a private subsidized school, to have completed high school as a maximum academic level reached, that your mother reached high school or university as the maximum complete academic level and reside in the Peñalolén area while you learn english. Surprisingly, age, to have had at least one subject taught 100% in english, the per capita income at home, to have lived in english speaking countries, to have studied for the english (TOEIC, etc) exam, to be self-taught and to have people to talk with in english with do not affect english level significantly.

## Contenido

Introducción .....	4
Origen e importancia .....	4
Revisión de literatura .....	4
Hipótesis del trabajo y objetivos .....	6
Materiales y métodos .....	7
Estadística descriptiva .....	14
Técnicas econométricas usadas .....	17
Resultados, discusión y conclusiones .....	19
Literatura citada .....	32

## Introducción

### Origen e importancia

El lenguaje, tanto escrito como hablado, es una herramienta que los seres humanos usan para comunicarse. En el mundo hay aproximadamente 6500 lenguajes hablados hoy en día. Es por esto que surge la necesidad de tener un lenguaje común para poder comunicarse con cualquier persona, cumpliendo este rol en la mayoría del mundo en la actualidad el idioma inglés. El inglés, si bien no es el lenguaje con más hablantes nativos (es el tercero, después del chino mandarín y el español), sí es el lenguaje con más hablantes totales (1.350 millones de personas aproximadamente) y es el lenguaje usado en todo tipo de eventos e instancias internacionales (negocios, seminarios, literatura académica, voluntariados, entre muchos otros) cuando personas que hablan diferentes lenguajes interactúan. Ahí radica la importancia de aprenderlo.

Actualmente el nivel de inglés de Chile es calificado como “moderado” por el índice EF, estando tercero en Latinoamérica después de Argentina y Costa Rica pero aún muy lejos de los países Escandinavos y Países Bajos, que son los países mejor evaluados según este índice. Pese a esto, del 2011 a la fecha el país ha ido subiendo su nivel desde “muy bajo” y pasando por “bajo”. Dentro de Chile, las zonas más pobladas (Metropolitana, Bio-bío y Valparaíso) fueron las mejores evaluadas, mientras que las zonas extremas (Magallanes y Atacama) las peores. También el segmento etario con mejor desempeño fueron los jóvenes de 21 a 30 años, lo que podría estar explicado por su acceso a la tecnología, videojuegos e internet (donde hay mucho material en inglés y acceso a practicar con personas de todo el mundo). Para saber cómo mejorar el nivel de inglés de los chilenos es necesario saber qué factores son más influyentes en este, lo que da origen a esta memoria.

### Revisión de literatura

A diferencia de la presente entrega, todas las investigaciones hechas en nuestro país no fueron aplicadas a nivel nacional (como población), si no sobre un grupo pequeño de personas (alumnos de colegios y universidad).

El año 2010 se realizó una investigación aplicada a alumnas del Liceo Santa María la Blanca de Valdivia (liceo femenino), y se enfocó en factores relacionados a la motivación e intereses. Fue un estudio basado en la percepción de alumnas y profesores basado en entrevistas. El estudio concluyó que incentivar el aprendizaje en las alumnas (tomando en cuenta sus intereses y habilidades), infraestructura y material, perseverancia y motivación (afectado principalmente por nivel socioeconómico – los más bajos no ven la necesidad de aprender inglés pues no lo ocupan – y autoestima) son los principales factores que afectan determinan el aprendizaje. Nótese que según la percepción de las alumnas, no aprenderían más si la asignatura de inglés fuese dictada 100% en este idioma (esta es una variable que se incluyó en la presente memoria).

En 2014 se realizó una investigación a un grupo de estudiantes de segundo medio de cuatro colegios de Chillán. Se midió el nivel de inglés mediante un test especial y se consideraron las variables explicativas nivel sociocultural, autoestima, ubicación geográfica y género. Se encontró que solamente nivel sociocultural y ubicación geográfica tenían relación con el nivel de inglés de los alumnos.

Finalmente (en el caso de Chile), en 2016 se realizó una investigación en la Universidad de Talca, donde se evaluó a estudiantes de nivel inicial de inglés de esta universidad. Se enfocó principalmente en si el uso de tecnología podía aumentar el aprendizaje de inglés. El nivel de inglés se evaluó mediante un test y se encontró que el establecimiento de origen (colegio) no tenía relación, que no había diferencias significativas entre las herramientas tecnológicas usadas en términos de resultados, que el profesor de inglés influía mucho en los resultados y que los alumnos bajaban su nivel al reemplazar al profesor por software.

En Colombia, en 2020 se realizó un estudio de percepción sobre el nivel de inglés en la universidad a estudiantes y profesores universitarios, los que dijeron que el currículum institucional, preparación de los profesores (nivel de conocimientos, actitud y motivar a los alumnos), papel de las políticas nacionales (principalmente asegurar acceso a educación pública), además de factores cognitivos (prestar atención a la clase) y afectivos como los factores más influyentes en el aprendizaje del inglés.

En México, un estudio de 2016 que tomó como muestra a a 897 estudiantes de primer año universitario y donde al igual que la presente memoria, se usó regresión múltiple, arrojaron como variables significativas a tiempo de estudio (horas de instrucción recibidas en inglés), tipo de escuela (particulares muestran mejor desempeño) y frecuencia con la que se lee en inglés.

Finalmente, en Perú, en 2020 se realizó una investigación sobre estudiantes de instituciones públicas, con una muestra de 284 estudiantes y aplicación de encuestas, además de un test de inglés y regresión múltiple, arrojó como variables que influyen en el nivel de inglés a motivación del estudiante, desempeño docente y estrategias de aprendizaje.

## Hipótesis del trabajo y objetivos

No hay una hipótesis en específico, si no que ver, dentro de los factores que constituyen la investigación y por ende la base de datos, ver qué factores influyen en el nivel de inglés de los chilenos, expresado en puntaje de pruebas TOEIC, TOEFL, Pearson academic (PTE) o IELTS, según sea el caso. Sin embargo, intuitivamente se espera que:

- No haya diferencia entre géneros masculino y femenino.
- Haya un efecto significativo de la edad, ya que se acumulan conocimientos.
- Mientras más privado el colegio, mejor puntaje.
- A mayor educación, mayor puntaje.
- Si la persona tuvo al menos una asignatura en inglés, que tenga mejor puntaje (aunque podría ser que no ya que se cree que mucha gente no contestó adecuadamente esa pregunta).
- A mayor ingreso por persona en el hogar en la época que aprendió más inglés, mayor puntaje.
- A mayor cantidad de tiempo vivido en país anglosajón antes de dar el examen, mayor puntaje.
- A mayor preparación específica para el examen, mayor puntaje.
- Los autodidactas tengan mejor desempeño (relacionado con motivación).
- A mayor educación de madre y padre, mayor puntaje (al menos hasta el nivel universitario, ya que está más relacionado con ingreso).
- Mayor cantidad de personas que le hablaban en inglés, mayor puntaje.
- Las comunas con mayor nivel socioeconómico obtengan mayor puntaje.

## Materiales y métodos

Origen de los datos: encuesta (vía google forms) difundida mayoritariamente en grupos de facebook y respondida por 125 personas, de las cuales se usarán 111 (significancia 5% y error 9.34%) ya que las demás no contestaron en la encuesta su puntaje en el examen de inglés correspondiente. Se usará una regresión lineal y distintas técnicas por explicar en el software Stata.

La encuesta tenía las siguientes preguntas:

- 1) Fecha de rendición del test (ya que ha habido diferencias en la escala de puntajes en el tiempo en algunos tests, por lo que quizás algunas respuestas podían no aplicar, aunque finalmente no hubieron problemas en ese aspecto).
- 2) Puntaje.
- 3) Test que rindió (alternativas IELTS, TOEIC, TOEFL y Pearson Academic PTE).
- 4) Género (alternativas: masculino y femenino).
- 5) Edad al rendir el test
- 6) Cursó la mayoría de su educación básica y media en (qué tipo de colegio). Alternativas: público, particular subvencionado, particular.
- 7) Nivel educativo al dar el examen. Alternativas: básica incompleta, básica, media, superior, magíster, doctorado o más.
- 8) Semestres de educación formal (incluyendo instituto de inglés o similares) donde la asignatura fue dictada 100% en inglés al momento de dar el test. Juzgando por las respuestas, se cree que muchas personas incluyeron sus asignaturas del ramo "inglés" aunque no hayan sido dictadas 100% en inglés, por lo que se decidió dejar como variable dummy.
- 9) Ingreso en su hogar en la época que aprendió más inglés, antes de dar el examen. Alternativas \$0 - \$500.000, \$500.001 - \$1.000.000, \$1.000.001 - \$1.500.000, \$1.500.001 - \$2.000.000, \$2.000.001 - \$2.500.000 y \$2.500.001 o más.
- 10) Cantidad de personas que componían su hogar en la época en que aprendió más inglés, antes de dar el examen. Los extremos del intervalo de 9) se dividieron por esta respuesta para cada observación y se sacó un promedio del intervalo para cada observación. Luego, se hicieron 6 clases con la técnica de clasificación de rupturas naturales de Jenks y se

clasificó cada uno de esos promedios en una clase (\$0 - \$149.999, \$150.000 - \$312.499, \$312.500 - \$499.999, \$500.000 - \$749.999, \$750.000 - \$1.124.999 y \$1.125.000 o más), generando variables dummy.

- 11) Meses vividos en país anglosajón antes de dar el examen.
- 12) Preparación específica para el examen (percepción), en escala de 1 a 10.
- 13) Percepción de la persona sobre su grado de autodidacta en la enseñanza del inglés, en escala de 1 a 10, antes de dar el examen. El objetivo de esta variable es representar de alguna manera la motivación por aprender el idioma.
- 14) Nivel educativo alcanzado por la madre. Alternativas: educación básica incompleta, básica completa, media completa, superior completa, magíster, doctorado o más. El objetivo de esta variable fue tratar de representar de alguna manera la genética del encuestado en cuanto a habilidades académicas.
- 15) Nivel educativo alcanzado por el padre. Alternativas: educación básica incompleta, básica completa, media completa, superior completa, magíster, doctorado o más. El objetivo de esta variable fue tratar de representar de alguna manera la genética del encuestado en cuanto a habilidades académicas.
- 16) Cantidad de personas que le hablaban en inglés (fuera de su educación formal – incluyendo instituto de inglés-) en la época que más aprendió el idioma, antes de dar el examen
- 17) Cantidad de personas que le hablaban en inglés (fuera de su educación formal – incluyendo instituto de inglés-) en su infancia y/o adolescencia.
- 18) Comuna de residencia en la época que aprendió más inglés antes de dar el examen. Para los que contestaron un lugar de algún país anglosajón, se agruparon todas esas respuestas como si hubiese sido la misma (todas como “país anglosajón”).

VARIABLES FINALES Y SU EXPLICACIÓN:

- Puntaje: puntaje obtenido por la persona en prueba TOEIC, TOEFL, Pearson academic o IELTS, según corresponda.
- Pearson: la persona rindió el examen Pearson Academic (PTE).
- TOEFL: la persona rindió el examen TOEFL.
- TOEIC: la persona rindió el examen TOEIC.

Cabe destacar que para estas variables dummy el grupo base fue el examen IELTS. Luego:

- Masculino: la persona es de género masculino (grupo base género femenino).

- Edad: edad de la persona al rendir el test.
- Partsub: la persona estudió la mayoría de su educación básica y media en un colegio particular subvencionado.
- Particular: la persona estudió la mayoría de su educación básica y media en un colegio particular.

El grupo base para las dos variables anteriores fue haber estudiado en un colegio público. Luego:

- Media: al rendir el examen, el mayor nivel académico alcanzado por la persona era haber terminado su educación media.
- Superior: al rendir el examen, el mayor nivel académico alcanzado por la persona era haber terminado su educación superior.
- Magíster: al rendir el examen, el mayor nivel académico alcanzado por la persona era haber terminado su magíster.
- Doctorado: al rendir el examen, el mayor nivel académico alcanzado por la persona era haber terminado su doctorado o más.

El grupo base para las cuatro variables anteriores fue haber alcanzado como máximo la educación básica terminada. Notar que pese a que el nivel “educación básica incompleta” estaba en las alternativas, nadie de la muestra la marcó. Luego:

- Cienporciento: la persona tuvo al menos un semestre de asignaturas dictados 100% en inglés en su educación formal (incluye instituto de inglés o similares), antes de dar el examen.
- Clasedos: en la época que aprendió más inglés (antes de rendir el examen), el ingreso por persona en el hogar del encuestado fue entre \$150.000 y \$312.499.
- Clasetres: en la época que aprendió más inglés (antes de rendir el examen), el ingreso por persona en el hogar del encuestado fue entre \$312.500 y \$499.999.
- Clasecuatro: en la época que aprendió más inglés (antes de rendir el examen), el ingreso por persona en el hogar del encuestado fue entre \$500.000 y \$749.999
- Clasecinco: en la época que aprendió más inglés (antes de rendir el examen), el ingreso por persona en el hogar del encuestado fue entre \$750.000 y \$1.124.999
- Claseseis: en la época que aprendió más inglés (antes de rendir el examen), el ingreso por persona en el hogar del encuestado fue de y \$1.125.000 o más

El grupo base para las cinco variables anteriores fue haber percibido menos de \$150.000 por persona. Luego:

- Anglosajón: cantidad de meses de residencia en país anglosajón de la persona, al momento de dar el examen.
- Específica: percepción de la persona, en escala de 1 a 10 (diez es más), sobre su propia preparación específica para dar el examen.
- Autodidacta: percepción de la persona, en escala de 1 a 10 (diez es más), sobre su grado de autodidacta en el aprendizaje del idioma inglés, antes de dar el examen.
- Mbásica: el máximo nivel educativo alcanzado por la madre del encuestado fue básica completa.
- Mmedia: el máximo nivel educativo alcanzado por la madre del encuestado fue media completa.
- Msuperior: el máximo nivel educativo alcanzado por la madre del encuestado fue superior completa.
- Mmagíster: el máximo nivel educativo alcanzado por la madre del encuestado fue magíster completo.
- Mdoctorado: el máximo nivel educativo alcanzado por la madre del encuestado fue doctorado o más.

El grupo base para las cinco variables dummy anteriores fue nivel educativo “básica incompleta” de la madre.

- Pbásica: el máximo nivel educativo alcanzado por el padre del encuestado fue básica completa.
- Pmedia: el máximo nivel educativo alcanzado por el padre del encuestado fue media completa.
- Psuperior: el máximo nivel educativo alcanzado por el padre del encuestado fue superior completa.
- Pmagíster: el máximo nivel educativo alcanzado por el padre del encuestado fue magíster completo.
- Pdoctorado: el máximo nivel educativo alcanzado por el padre del encuestado fue doctorado o más.

El grupo base para las cinco variables dummy anteriores fue nivel educativo “básica incompleta” del padre. Luego:

- Hablaban: cantidad de personas que le hablaban al encuestado regularmente en inglés en la época que aprendió más inglés, antes de dar el examen.
- Infanciadolescencia: cantidad de personas que le hablaban al encuestado regularmente en inglés en su infancia y/o adolescencia.
- Arica: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Arica.
- Chillán: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Chillán.
- Concepción: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Concepción.
- Conchalí: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Conchalí.
- Copiapó: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Copiapó.
- Coquimbo: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Coquimbo.
- Coronel: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Coronel.
- Curicó: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Curicó.
- Elbosque: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en la comuna de El Bosque.
- Elmonte: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en la comuna de El Monte.
- Hualpén: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Hualpén.
- Iquique: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Iquique.
- Laflorida: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en La Florida.

- Lagranja: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en la comuna de La Granja.
- Lareina: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en la comuna de La Reina.
- Laserena: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en La Serena.
- Lascondes: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Las Condes.
- Linares: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Linares.
- Lobarnechea: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Lo Barnechea.
- Loprado: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Lo Prado.
- Machalí: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Machalí.
- Magallanes: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Magallanes.
- Maipú: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Maipú.
- Ñuñoa: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Ñuñoa.
- Oslo: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Oslo, Noruega.
- Padrehurtado: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en la comuna de Padre Hurtado.
- Paísesanglosajón: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en un país anglosajón.
- Peñalolen: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Peñalolén.
- Petorca: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Petorca.

- Providencia: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Providencia.
- Puentealto: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Puente Alto.
- Puertomontt: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Puerto Montt.
- Puntaarenas: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Punta Arenas.
- Quilicura: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Quilicura.
- Quilpué: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Quilpué.
- Quintanormal: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Quinta Normal.
- Renca: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Renca.
- Sanbernardo: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en San Bernardo.
- Santacruz: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Santa Cruz.
- Santiago: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en la comuna de Santiago.
- Talca: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Talca.
- Valparaíso: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en la comuna de Valparaíso.
- Villaalemana: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Villa Alemana.
- Viñadelmar: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Viña del Mar.
- Vitacura: en la época en que el encuestado aprendió más inglés (antes de dar el examen), residía en Vitacura.

Para las 45 variables dummy anteriores, el grupo base fue la comuna de Ancud.

Estadística descriptiva

Variable	Obs	Mean	Std. Dev.	Min	Max
puntaje	111	170.6306	276.5545	4.5	990
pearson	111	.0540541	.2271498	0	1
toefl	111	.3873874	.4893627	0	1
toeic	111	.1981982	.4004502	0	1
masculino	111	.3963964	.4913669	0	1
edad	111	27.27027	5.344404	17	50
partsub	111	.4054054	.493197	0	1
particular	111	.3243243	.4702449	0	1
media	111	.2162162	.4135304	0	1
superior	111	.6846847	.4667486	0	1
Magister	111	.0810811	.2741975	0	1
doctorado	111	.009009	.0949158	0	1
cienporcie-o	111	.7657658	.42544	0	1
clasesdos	109	.2293578	.4223617	0	1
clasetres	109	.2844037	.4532137	0	1
clasecuatro	109	.2201835	.4162842	0	1
clasecinco	109	.1100917	.3144498	0	1
claseseis	109	.0366972	.1888859	0	1
Anglosajón	106	6.471698	16.72535	0	108
Especifica	111	5.297297	2.752171	1	10
autodidacta	111	7.324324	2.04567	1	10
Mbásica	111	.0540541	.2271498	0	1
mmedia	111	.4594595	.5006139	0	1
msuperior	111	.3873874	.4893627	0	1
Mmagister	111	.0630631	.2441787	0	1
mdoctorado	111	.009009	.0949158	0	1
Pbásica	111	.027027	.1628976	0	1
pmedia	111	.3063063	.463049	0	1
psuperior	111	.4774775	.5017578	0	1
Pmagister	111	.1081081	.3119251	0	1
pdoctorado	111	.009009	.0949158	0	1
hablaban	103	3.339806	8.631399	0	50
infanciado-a	109	.266055	.7154667	0	4
arica	111	.009009	.0949158	0	1
Chillán	111	.009009	.0949158	0	1
Concepción	111	.036036	.1872251	0	1
Conchalí	111	.009009	.0949158	0	1
Copiapó	111	.009009	.0949158	0	1
coquimbo	111	.009009	.0949158	0	1
coronel	111	.009009	.0949158	0	1
Curicó	111	.009009	.0949158	0	1
elbosque	111	.018018	.1336197	0	1
elmonte	111	.009009	.0949158	0	1
Hualpén	111	.009009	.0949158	0	1
iquique	111	.018018	.1336197	0	1
laflorida	111	.036036	.1872251	0	1
lagranja	111	.018018	.1336197	0	1
lareina	111	.027027	.1628976	0	1
laserena	111	.009009	.0949158	0	1
lascondes	111	.027027	.1628976	0	1
linares	111	.009009	.0949158	0	1
lobarnechea	111	.009009	.0949158	0	1
loprado	111	.009009	.0949158	0	1
Machali	111	.009009	.0949158	0	1
magallanes	111	.009009	.0949158	0	1
Maipú	111	.018018	.1336197	0	1
Rufoa	111	.018018	.1336197	0	1
oslo	111	.009009	.0949158	0	1
padrehurtado	111	.009009	.0949158	0	1
Faisanglos-n	111	.1351351	.3434183	0	1
Peñalolén	111	.036036	.1872251	0	1
petorca	111	.009009	.0949158	0	1
providencia	111	.036036	.1872251	0	1
puntealto	111	.036036	.1872251	0	1
puertomontt	111	.027027	.1628976	0	1
puntaarenas	111	.009009	.0949158	0	1
quilicura	111	.009009	.0949158	0	1
Quilpué	111	.018018	.1336197	0	1
quintanormal	111	.018018	.1336197	0	1
renca	111	.027027	.1628976	0	1
sanbernardo	111	.018018	.1336197	0	1
santacruz	111	.009009	.0949158	0	1
santiago	111	.0990991	.3001501	0	1
talca	111	.009009	.0949158	0	1
Valparaíso	111	.0540541	.2271498	0	1
villaalemana	111	.009009	.0949158	0	1
Viñadelmar	111	.045045	.2083436	0	1
vitacura	111	.018018	.1336197	0	1

Algunos aspectos a destacar de esta parte de la estadística descriptiva respecto a variables no binarias:

La edad que tenían los encuestados al dar el test varía entre 17 y 50 años, siendo el promedio de edad 27 años.

- Los meses que los encuestados residieron en países anglosajones al dar el test varía entre 0 y 108 meses (11 años), siendo la media 6.47 meses y el percentil 75 seis meses.

#### Anglosajón

	Percentiles	Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	106
25%	0	0	Sum of Wgt.	106
50%	0		Mean	6.471698
		Largest	Std. Dev.	16.72535
75%	6	36		
90%	18	60	Variance	279.7373
95%	30	102	Skewness	4.384025
99%	102	108	Kurtosis	24.5869

- El promedio de percepción de preparación específica para el examen fue de 5.29, variando entre 1 y 10.
- El promedio de percepción de autodidacta fue de 7.3, variando entre 1 y 10.
- La cantidad de personas que le hablaban en inglés de manera recurrente en la época que aprendió más inglés antes de dar el examen, varió entre 0 y 50, siendo el promedio 3.33 y el percentil 75 tres personas.

#### Hablaban

	Percentiles	Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	103
25%	0	0	Sum of Wgt.	103
50%	0		Mean	3.339806
		Largest	Std. Dev.	8.631399
75%	3	30		
90%	10	30	Variance	74.50105
95%	15	50	Skewness	3.904083
99%	50	50	Kurtosis	19.23473

- La cantidad de personas que le hablaban en inglés de manera recurrente en su infancia y/o adolescencia varió entre 0 y 4, siendo el promedio 0.26.

Respecto de variables binarias:

- Los tests más rendidos en la muestra (75% aproximadamente) fueron TOEFL e IELTS.
- El 60% aproximadamente de la muestra fueron mujeres.
- El tipo de colegio más repetido en la muestra fue particular subvencionado, aunque con poca diferencia respecto de los otros dos.
- El 68% de la muestra tenía título universitario al dar el test respectivo.
- Los padres de los encuestados en su mayoría (78% para padres y 85% para madres) completaron educación media o universitaria como máximo nivel.
- 48% de los encuestados residía en la Región Metropolitana cuando aprendió más inglés, 38% en regiones y 14% en el extranjero (casi todos en países anglosajones).

Y además:

- Dado que algunas preguntas fueron dejadas en blanco o mal contestadas, dependiendo de la variable la muestra fluctúa entre 103 y 111 personas.

### Técnicas econométricas usadas

La base estadística es la regresión lineal múltiple, la que consiste en aproximar observaciones (en este caso datos de corte transversal, que representan un momento del tiempo) mediante una línea. Se explica una variable Y (dependiente, explicada, respuesta, predicha o regresando) por variables X (independientes, explicativas, control, predictoras, o regresores). También hay un término “u” o error o perturbación, donde se representan factores distintos a las X que afectan a Y. La estimación de la regresión lineal múltiple se realizó mediante un método llamado mínimos cuadrados ordinarios (MCO), donde se ajusta una línea a través de los puntos de la muestra de manera que la suma de cuadrados de los residuales ( $\text{residual} = Y \text{ observado} - Y \text{ estimado}$ ) sea lo más pequeña posible.

Cuando se habla del  $R^2$ , es un índice de “bondad de ajuste”, que habla de en qué porcentaje las variables X explican a la variable Y.

Los indicadores para detectar observaciones influyentes (también llamados outliers, que influyen en la muestra de tal manera que distorsionan la inferencia) usados son: residuos estudentizados (residuales divididos por sus errores estándar estimados), valores leverage (mide cuán lejos una observación está de otras basándose en los niveles de las variables independientes), DFFITS (medida de cuánto una observación ha afectado a su valor ajustado), D de Cook (medida de impacto agregado de cada observación en el grupo de coeficientes de regresión, así como en el grupo de valores ajustados) y COVRATIO (medida del impacto de cada observación en las varianzas y errores estándares de los coeficientes de regresión y sus covarianzas).

El VIF (variance inflation factor) es una medida de colinealidad, donde alrededor de 10 hacia arriba se considera que hay colinealidad alta. La colinealidad ocurre cuando más de una variable independiente explica lo mismo de la variable dependiente, y tiene que ver con el grado de asociación lineal entre variables.

La heterocedasticidad (que es un problema, que rompe con los supuestos de Gauss-Markov) ocurre cuando la varianza de los residuales para cierta variable no es constante. En este trabajo se usaron los errores estándares robustos (o no sesgados) para corregirla.

La selección de modelos, es un método por el cual se prueban todos los posibles modelos dadas las variables existentes, para determinar cuál es el mejor, dado un nivel de significancia para incluir o remover una variable del modelo. El método usado en el presente trabajo es “stepwise”, es decir, comenzando con el modelo con todas las variables e ir eliminando éstas y comenzar con el modelo vacío e ir adicionando variables. Ambas a la vez.

Finalmente, la normalización del modelo, consiste en “normalizar” (restar el promedio y dividir por la desviación estándar de la variable) cada observación, donde da un modelo que al correrse, da coeficientes que determinan cuánto cada variable (no en magnitud, si no que para ver qué variables influyen más o menos en la variable dependiente) influye en la variable dependiente.

## Resultados, discusión y conclusiones

Hecha la regresión queda:

```

. regress puntaje pearson toefl toefc masculino edad partsub particular media superior Magister doctorado cienporciento clasedos clasetres cl
> asecatro clasecinco claseeis Anglosajón Especifica autodidacta Mbásica mmedia msuperior Mmagister mdoctorado Pbásica pmedia psuperior Pma
> gister pdoctorado hablaban infanciadolescencia arica Chillán Concepción Conchali Copiapó coquimbo coronel Curicó elbosque elmonte Hualpén i
> quique laflorida lagranja lareina laserena lascondes linares lobarnechea loprado Machali magallanes Maipú Ñuñoa oslo padrehurtado Paisangl
> sajón Peñalolén petorca providencia puentealto puertomont puntaarenas quilicura Quilpué quintanormal renca sanbernardo santacruz santiago
> talca Valparaiso villaalemana Viñadelmar vitacura
note: mdoctorado omitted because of collinearity
note: pdoctorado omitted because of collinearity
note: Chillán omitted because of collinearity
note: coronel omitted because of collinearity

```

Source	SS	df	MS	Number of obs	=	98
Model	7737780.16	73	105996.988	F(73, 24)	=	6.22
Residual	408834.068	24	17034.7528	Prob > F	=	0.0000
				R-squared	=	0.9498
				Adj R-squared	=	0.7972
Total	8146614.23	97	83985.7137	Root MSE	=	130.52

puntaje	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
pearson	224.6281	134.8988	1.67	0.109	-53.78932 503.0455
toefl	104.8871	64.58815	1.62	0.117	-28.41631 238.1904
toefc	714.6109	95.78343	7.46	0.000	516.9236 912.2981
masculino	96.28507	52.44045	1.84	0.079	-11.94669 204.5168
edad	.0017624	7.799969	0.00	1.000	-16.09658 16.10011
partsub	-132.7467	70.88881	-1.87	0.073	-279.054 13.5606
particular	-32.05959	66.21731	-0.48	0.633	-168.7254 104.6062
media	263.7244	256.8207	1.03	0.315	-266.3275 793.7763
superior	262.7015	278.64	0.94	0.355	-312.3832 837.7862
Magister	323.6493	291.1591	1.11	0.277	-277.2735 924.5721
doctorado	218.0132	399.9122	0.55	0.591	-607.3649 1043.391
cienporciento	-26.09827	58.79524	-0.44	0.661	-147.4457 95.24915
clasedos	119.0727	82.22296	1.45	0.161	-50.62716 288.7725
clasetres	244.6637	105.5927	2.32	0.029	26.73114 462.5962
clasecuatro	171.2424	88.518	1.93	0.065	-11.44972 353.9346
clasecinco	267.1877	92.35784	2.89	0.008	76.5705 457.8049
claseeis	249.85	129.695	1.93	0.066	-17.80661 517.5066
Anglosajón	.2296869	2.108601	0.11	0.914	-4.122252 4.581626
Especifica	-7.053291	11.63162	-0.61	0.550	-31.05978 16.95319
autodidacta	-10.39539	21.00542	-0.49	0.625	-53.74844 32.95767
Mbásica	59.18994	189.7023	0.31	0.758	-332.3364 450.7163
mmedia	261.9668	187.5891	1.40	0.175	-125.1981 649.1316
msuperior	261.3499	184.3028	1.42	0.169	-119.0325 641.7323
Mmagister	207.3462	210.6248	0.98	0.335	-227.3621 642.0545
mdoctorado	0	(omitted)			
Pbásica	288.4267	245.3354	1.18	0.251	-217.9208 794.7742
pmedia	60.48542	160.1806	0.38	0.709	-270.1112 391.082
psuperior	-45.46901	142.1638	-0.32	0.752	-338.8808 247.9427
Pmagister	186.9731	205.0753	0.91	0.371	-236.2815 610.2277
pdoctorado	0	(omitted)			
hablaban	-3.788119	3.814674	-0.99	0.331	-11.66122 4.084982
infanciadolescencia	-50.4352	48.56498	-1.04	0.309	-150.6684 49.798
arica	54.28071	242.7583	0.22	0.825	-446.7479 555.3093
Chillán	0	(omitted)			
Concepción	9.696424	212.3183	0.05	0.964	-428.5069 447.8998
Conchali	-96.70935	281.5372	-0.34	0.734	-677.7736 484.3549
Copiapó	-217.9847	271.9337	-0.80	0.431	-779.2282 343.2588
coquimbo	-168.6809	249.1428	-0.68	0.505	-682.8864 345.5245
coronel	0	(omitted)			
Curicó	-35.05284	292.6767	-0.12	0.906	-639.1079 569.0023
elbosque	-48.88835	252.0278	-0.19	0.848	-569.0481 471.2714
elmonte	-243.1511	349.2894	-0.70	0.493	-964.0491 477.7468
Hualpén	38.69668	234.4939	0.17	0.870	-445.275 522.6683
iquique	177.1233	217.9641	0.81	0.424	-272.7325 626.9791
laflorida	-1.882426	228.9659	-0.01	0.994	-474.4448 470.6799
lagranja	-81.83765	275.5232	-0.30	0.769	-650.4896 486.8143
lareina	-21.13065	212.1058	-0.10	0.921	-458.8956 416.6343
laserena	-24.77495	243.5933	-0.10	0.920	-527.5269 477.9769
lascondes	201.7796	236.511	0.85	0.402	-286.355 689.9142
linares	52.6311	244.3756	0.22	0.831	-451.7353 556.9975
lobarnechea	-158.3964	345.6516	-0.46	0.651	-871.7863 554.9934
loprado	48.02513	344.8987	0.14	0.890	-663.8108 759.8611
Machali	-34.33516	244.1953	-0.14	0.889	-538.3296 469.6593
magallanes	164.9886	328.7156	0.50	0.620	-513.4471 843.4244
Maipú	38.00607	246.7732	0.15	0.879	-471.3088 547.3209
Ñuñoa	30.14762	232.321	0.13	0.898	-449.3393 509.6345
oslo	-31.03137	303.7619	-0.10	0.919	-657.9652 595.9025
padrehurtado	-49.05083	308.8185	-0.16	0.875	-686.4208 588.3192
Paisanglosajón	-4.1212	229.7162	-0.02	0.986	-478.2322 469.9898
Peñalolén	40.56254	217.8353	0.19	0.854	-409.0275 490.1526
petorca	286.5193	263.93	1.09	0.288	-258.2054 831.2441
providencia	-97.15843	229.7145	-0.42	0.676	-571.2658 376.9489
puentealto	16.45199	214.7666	0.08	0.940	-426.8044 459.7084
puertomont	-122.8008	265.8884	-0.46	0.648	-671.5675 425.9658
puntaarenas	-224.0003	256.8297	-0.87	0.392	-754.0708 306.0702
quilicura	19.85096	240.9944	0.08	0.935	-477.5369 517.2389
Quilpué	-485.4197	252.9853	-1.92	0.067	-1007.556 36.71621
quintanormal	143.6688	222.0733	0.65	0.524	-314.668 602.0056
renca	52.96048	223.5071	0.24	0.815	-408.3354 514.2564
sanbernardo	-6.668938	229.8266	-0.03	0.977	-481.0076 467.6698
santacruz	-142.9056	272.2679	-0.52	0.604	-704.8389 419.0276
santiago	-150.1451	227.7095	-0.66	0.516	-620.1145 319.8243
taica	-272.6447	281.9861	-0.97	0.343	-854.6354 309.346
Valparaiso	-164.6428	244.8843	-0.67	0.508	-670.0592 340.7736
villaalemana	-105.2044	248.8203	-0.42	0.676	-618.7443 408.3356
Viñadelmar	-211.2389	238.7846	-0.88	0.385	-704.0662 281.5883
vitacura	-80.85094	281.2087	-0.29	0.776	-661.2372 499.5353
_cons	-493.1957	363.9258	-1.36	0.188	-1244.302 257.9103

Podemos ver que el software indica que las variables “Mdoctorado”, “Pdoctorado”, “Chillán” y “Coronel” fueron desechadas por perfecta colinealidad. También podemos ver que el modelo es significativo con un 95% de confianza (P value menor a 5%) y que de acuerdo a la significancia local, “Clasecinco”, “Clasetres” y “TOEIC” son las únicas variables significativas. Además, el R cuadrado es de casi un 95%.

Y hecha la selección de modelos con una significancia de 15%, queda que:

Source	SS	df	MS	Number of obs	=	98
				F(21, 76)	=	42.31
Model	7504706.11	21	357366.958	Prob > F	=	0.0000
Residual	641908.118	76	8446.15945	R-squared	=	0.9212
				Adj R-squared	=	0.8994
Total	8146614.23	97	83985.7137	Root MSE	=	91.903

puntaje	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
pearson	141.2431	49.33124	2.86	0.005	42.99141	239.4948
toefl	99.82237	24.8367	4.02	0.000	50.35579	149.289
toeic	648.2799	31.46281	20.60	0.000	585.6163	710.9435
masculino	64.0563	20.76734	3.08	0.003	22.69455	105.418
iquique	180.8085	71.71379	2.52	0.014	37.97808	323.6389
partsub	-48.55422	22.24508	-2.18	0.032	-92.85913	-4.249296
loprado	260.6924	106.2851	2.45	0.016	49.00719	472.3776
media	59.29134	26.06558	2.27	0.026	7.377235	111.2054
Quilpué	-250.2034	74.12933	-3.38	0.001	-397.8448	-102.5621
Peñalolén	114.2676	48.86871	2.34	0.022	16.93709	211.598
msuperior	116.7301	31.52234	3.70	0.000	53.94795	179.5123
mmedia	119.5119	31.18607	3.83	0.000	57.39945	181.6243
clasedos	155.7264	40.89733	3.81	0.000	74.2723	237.1805
clasetres	182.8224	42.22915	4.33	0.000	98.71576	266.9291
clasecuatro	174.2417	43.44665	4.01	0.000	87.71017	260.7732
clasecinco	182.7337	48.95485	3.73	0.000	85.2317	280.2358
claseseis	160.3854	60.00753	2.67	0.009	40.87009	279.9008
psuperior	-49.5125	21.81479	-2.27	0.026	-92.96042	-6.064588
lascondes	240.6501	58.08666	4.14	0.000	124.9605	356.3397
quintanormal	192.4199	72.92422	2.64	0.010	47.17875	337.6611
petorca	337.4131	102.4847	3.29	0.002	133.2972	541.5291
_cons	-271.1945	53.7446	-5.05	0.000	-378.2361	-164.1529

Vemos que el modelo es significativo, el r cuadrado ajustado subió de 80% aproximadamente a 90% aproximadamente y quedaron 21 variables explicativas, todas ahora significativas al 5%. Sin embargo, al correr el modelo solamente con esas variables, “psuperior” salió no significativa:

Source	SS	df	MS	Number of obs	=	109
				F(21, 87)	=	42.02
Model	7629219.02	21	363296.144	Prob > F	=	0.0000
Residual	752269.746	87	8646.77869	R-squared	=	0.9102
				Adj R-squared	=	0.8886
Total	8381488.77	108	77606.3775	Root MSE	=	92.988

puntaje	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
pearson	132.8421	49.06241	2.71	0.008	35.32522	230.3589
toefl	94.62472	23.26836	4.07	0.000	48.37633	140.8731
toeic	623.0447	30.39592	20.50	0.000	562.6295	683.4599
masculino	57.53949	19.93161	2.89	0.005	17.92326	97.15571
partsub	-45.69328	20.92055	-2.18	0.032	-87.27514	-4.111434
media	60.07623	25.45499	2.36	0.021	9.481685	110.6708
clasedos	149.8081	37.08828	4.04	0.000	76.0911	223.525
clasetres	165.5096	38.13988	4.34	0.000	89.70242	241.3167
clasecuatro	161.5765	39.72755	4.07	0.000	82.6137	240.5393
clasecinco	192.5494	43.28788	4.45	0.000	106.5101	278.5888
claseseis	154.5641	58.30327	2.65	0.010	38.68003	270.4481
mmedia	133.5582	28.96118	4.61	0.000	75.99475	191.1217
msuperior	125.2464	29.54269	4.24	0.000	66.52713	183.9657
psuperior	-27.8999	20.23526	-1.38	0.171	-68.11966	12.31986
iquique	176.1392	71.7455	2.46	0.016	33.53721	318.7411
lascondes	252.4918	58.1959	4.34	0.000	136.8212	368.1625
loprado	265.5764	104.761	2.54	0.013	57.3525	473.8003
Peñalolén	120.1257	49.27258	2.44	0.017	22.19109	218.0603
petorca	331.8734	101.6546	3.26	0.002	129.8238	533.923
Quilpué	-225.0592	73.91864	-3.04	0.003	-371.9805	-78.13786
quintanormal	207.8011	72.62583	2.86	0.005	63.44939	352.1528
_cons	-272.4226	47.21984	-5.77	0.000	-366.2771	-178.5681

Es por esto que se corrió una nueva selección de modelos con estas variables, donde el software eliminó la variable “psuperior”:

Source	SS	df	MS	Number of obs	=	109
Model	7612781.27	20	380639.064	F(20, 88)	=	43.57
Residual	768707.494	88	8735.31243	Prob > F	=	0.0000
				R-squared	=	0.9083
				Adj R-squared	=	0.8874
Total	8381488.77	108	77606.3775	Root MSE	=	93.463

puntaje	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
pearson	128.5098	49.21171	2.61	0.011	30.71191	226.3078
toefl	91.02266	23.2393	3.92	0.000	44.83944	137.2059
toeic	619.8928	30.46461	20.35	0.000	559.3508	680.4348
masculino	56.28742	20.01258	2.81	0.006	16.51661	96.05822
partsub	-44.28878	21.00243	-2.11	0.038	-86.0267	-2.550857
media	57.23155	25.50079	2.24	0.027	6.554082	107.909
clasesdos	150.3623	37.27547	4.03	0.000	76.28513	224.4395
clasetres	159.7295	38.10237	4.19	0.000	84.00907	235.45
clasecuatro	160.6682	39.92492	4.02	0.000	81.32585	240.0106
clasecinco	190.9045	43.4924	4.39	0.000	104.4724	277.3365
claseseseis	164.9824	58.10675	2.84	0.006	49.50745	280.4574
mmedia	137.9445	28.93291	4.77	0.000	80.44642	195.4426
msuperior	123.4448	29.66449	4.16	0.000	64.49286	182.3967
quintanormal	217.6895	72.63989	3.00	0.004	73.33299	362.046
iquique	171.2329	72.02312	2.38	0.020	28.10213	314.3638
lascondes	251.3606	58.48726	4.30	0.000	135.1295	367.5917
loprado	276.18	105.0119	2.63	0.010	67.491	484.8691
Peñalolén	120.497	49.52345	2.43	0.017	22.07951	218.9144
petorca	313.3427	101.2768	3.09	0.003	112.0764	514.6091
Quilpué	-203.7408	72.65257	-2.80	0.006	-348.1225	-59.35909
_cons	-283.1787	46.80878	-6.05	0.000	-376.2013	-190.1561

Se ve que este modelo es significativo, el r2 ajustado es de 88,74% y quedaron 20 variables explicativas, todas significativas al 5%. Luego, corremos el modelo sólo con estas variables:

Source	SS	df	MS	Number of obs	=	109
Model	7612781.27	20	380639.064	F(20, 88)	=	43.57
Residual	768707.494	88	8735.31243	Prob > F	=	0.0000
				R-squared	=	0.9083
				Adj R-squared	=	0.8874
Total	8381488.77	108	77606.3775	Root MSE	=	93.463

puntaje	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
pearson	128.5098	49.21171	2.61	0.011	30.71191	226.3078
toefl	91.02266	23.2393	3.92	0.000	44.83944	137.2059
toeic	619.8928	30.46461	20.35	0.000	559.3508	680.4348
masculino	56.28742	20.01258	2.81	0.006	16.51661	96.05822
partsub	-44.28878	21.00243	-2.11	0.038	-86.0267	-2.550857
media	57.23155	25.50079	2.24	0.027	6.554082	107.909
clasedos	150.3623	37.27547	4.03	0.000	76.28513	224.4395
clasetres	159.7295	38.10237	4.19	0.000	84.00907	235.45
clasecuatro	160.6682	39.92492	4.02	0.000	81.32585	240.0106
clasecinco	190.9045	43.4924	4.39	0.000	104.4724	277.3365
claseseis	164.9824	58.10675	2.84	0.006	49.50745	280.4574
mmedia	137.9445	28.93291	4.77	0.000	80.44642	195.4426
msuperior	123.4448	29.66449	4.16	0.000	64.49286	182.3967
iquique	171.2329	72.02312	2.38	0.020	28.10213	314.3638
lascondes	251.3606	58.48726	4.30	0.000	135.1295	367.5917
loprado	276.18	105.0119	2.63	0.010	67.491	484.8691
Peñalolén	120.497	49.52345	2.43	0.017	22.07951	218.9144
petorca	313.3427	101.2768	3.09	0.003	112.0764	514.6091
Quilpué	-203.7408	72.65257	-2.80	0.006	-348.1225	-59.35909
quintanormal	217.6895	72.63989	3.00	0.004	73.33299	362.046
_cons	-283.1787	46.80878	-6.05	0.000	-376.2013	-190.1561

Se ve que el modelo se mantuvo significativo, con el mismo r2 ajustado y con todas las variables significativas al 5%, por lo que este modelo es el que se usará ahora para eliminar posibles observaciones influyentes. Para esto, se usarán los indicadores “leverage values”, “studentized residuals”, “Cook’s D”, “COVRATIO” y “DFITS”. Como resultado, se eliminaron las observaciones 4, 11, 19, 21, 27, 33, 41, 43, 44, 46, 66, 79, 90, 91, 92, 94, 95, 97, 98, 99, 102, 103, 106, 107, 109 y 110. Con esto, quedaron un total de 85 observaciones de las 111 originales.

Luego se verá la colinealidad por VIF. Se corre el modelo, y el software elimina por colinealidad las variables correspondientes a las comunas de Iquique, Las Condes, Lo Prado, Petorca, Quilpué y Quinta Normal. Al hacer la prueba de VIF:

Variable	VIF	1/VIF
clasetres	7.26	0.137678
clasecuatro	6.48	0.154385
clasedos	6.02	0.166056
clasecinco	4.26	0.234478
mmedia	3.67	0.272676
msuperior	3.56	0.280525
claseseis	1.52	0.659984
toefl	1.39	0.717463
media	1.37	0.732212
toeic	1.32	0.755648
partsub	1.27	0.786595
masculino	1.24	0.805031
pearson	1.13	0.884863
Peñalolén	1.08	0.925838
Mean VIF	2.97	

Se ve que ninguna variable tiene VIF alto.

Luego, se verá la posible presencia de heterocedasticidad.

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
```

```
Ho: Constant variance
```

```
Variables: pearson toefl toeic masculino partsub media clasedos clasetres clasecuatro clasecinco claseseis mmedia msuperior  
iquique lascondes lo Prado Peñalolén petorca Quilpué quintanormal
```

```
chi2(14) = 174.55
```

```
Prob > chi2 = 0.0000
```

```
. estat hetttest, fstat
```

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
```

```
Ho: Constant variance
```

```
Variables: fitted values of puntaje
```

```
F(1 , 81) = 71.68
```

```
Prob > F = 0.0000
```

Se ve que por ambas pruebas (Breusch-Pagan y White) el modelo es heterocedástico. Se usará el método de los errores estándares robustos para corregir heterocedasticidad.

puntaje	Robust		t	P> t	[95% Conf. Interval]	
	Coef.	Std. Err.				
pearson	24.33014	12.00767	2.03	0.047	.369204	48.29108
toefl	74.33876	6.43284	11.56	0.000	61.50222	87.17529
toeic	609.1255	31.09049	19.59	0.000	547.0854	671.1656
masculino	30.46849	9.438152	3.23	0.002	11.63495	49.30203
partsub	-17.92809	9.491933	-1.89	0.063	-36.86895	1.012769
media	18.70132	11.37204	1.64	0.105	-3.991228	41.39387
clasedos	62.65146	31.65337	1.98	0.052	-.5118589	125.8148
clasetres	41.58869	28.70465	1.45	0.152	-15.69054	98.86793
clasecuatro	48.34822	30.21053	1.60	0.114	-11.93595	108.6324
clasecinco	37.8067	30.91576	1.22	0.226	-23.88475	99.49814
claseseis	30.32068	32.02295	0.95	0.347	-33.58011	94.22147
mmedia	20.44067	11.17693	1.83	0.072	-1.862539	42.74388
msuperior	29.46159	10.55602	2.79	0.007	8.397373	50.52581
iunque	0	(omitted)				
lascondes	0	(omitted)				
loprado	0	(omitted)				
Peñalolén	28.50172	9.362355	3.04	0.003	9.819432	47.18401
petorca	0	(omitted)				
Quilpué	0	(omitted)				
quintanormal	0	(omitted)				
_cons	-61.96268	31.84738	-1.95	0.056	-125.5131	1.587785

Luego, se verá qué variables influyen más en el nivel de inglés de los chilenos aproximado por el puntaje en los respectivos tests. Se estandariza la base de datos (restando cada observación por el promedio y dividiendo por la desviación estándar de la respectiva variable) y se corre la regresión, resultando:

Source	SS	df	MS	Number of obs	=	85
Model	75.7211644	14	5.4086546	F(14, 70)	=	168.21
Residual	2.25085864	70	.032155123	Prob > F	=	0.0000
Total	77.9720231	84	.92823837	R-squared	=	0.9711
				Adj R-squared	=	0.9654
				Root MSE	=	.17932

puntaje	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
pearson	.0154566	.0218162	0.71	0.481	-.0280544	.0589677
toefl	.2086669	.0230974	9.03	0.000	.1626006	.2547331
toeic	.972501	.0225193	43.19	0.000	.9275877	1.017414
masculino	.0762255	.0214791	3.55	0.001	.0333868	.1190641
partsub	-.0513942	.0218234	-2.36	0.021	-.0949196	-.0078688
media	.0370591	.0231969	1.60	0.115	-.0092056	.0833238
clasedos	.1113531	.0394391	2.82	0.006	.0326942	.1900119
clasetres	.0712554	.0442465	1.61	0.112	-.0169914	.1595023
clasecuatro	.0808755	.04181	1.93	0.057	-.0025118	.1642629
clasecinco	.0440733	.0353171	1.25	0.216	-.0263645	.1145111
claseseis	.0094974	.0230027	0.41	0.681	-.03638	.0553747
mmedia	.0573959	.0379038	1.51	0.134	-.0182008	.1329926
msuperior	.0757368	.0371549	2.04	0.045	.0016336	.1498399
Peñalolén	.0168542	.0203362	0.83	0.410	-.0237052	.0574135
_cons	-.0602101	.0195511	-3.08	0.003	-.0992035	-.0212166

Se ve que las variables que influyen de manera positiva, en orden decreciente, son:

- 1) Clasedos (es decir que si el ingreso aproximado por persona en el hogar en la época que aprendió más inglés va entre \$150.000 y \$312.499 es mejor para el puntaje).
- 2) Clasecuatro (ingreso aproximado por persona en el hogar en la época que aprendió más inglés va entre \$500.000 y \$749.999).
- 3) Masculino (encuestado de género masculino).
- 4) Msuperior (educación de la madre llegó hasta nivel superior completo)
- 5) Clasetres (ingreso aproximado por persona en el hogar en la época que aprendió más inglés va entre \$312.500 y \$499.999).
- 6) Mmedia (nivel educativo máximo de la madre del encuestado es educación media).

- 7) Clasecinco (ingreso aproximado por persona en el hogar en la época que aprendió más inglés va entre \$750.000 y \$1.124.999).
- 8) Media (al momento de dar el examen, encuestado llegó hasta enseñanza media completa).
- 9) Peñalolén (encuestado vivía en Peñalolén en la época que aprendió más inglés).
- 10) Claseseis (ingreso aproximado por persona en el hogar en la época que aprendió más inglés es de \$1.125.000 o más).

Solamente la variable "Partsub" (encuestado estudió la mayoría de su vida escolar en colegio particular subvencionado) afecta de manera negativa (contrario a lo que se pueda pensar, ya que la base es colegio público).

De la normalización, se puede concluir que:

- 1) No necesariamente el nivel socioeconómico (o su aproximación en esta encuesta) afecta el nivel de inglés (ya que clasedos > clasecuatro > clasetres > clasecinco > claseseis)
- 2) El nivel educativo de la madre afecta en el orden esperado (mayor nivel, mayor puntaje).

La regresión final, luego de todos los procedimientos indicados, queda (con 85 observaciones, correspondientes a 95% de confianza y 10,65% de error):

Source	SS	df	MS	Number of obs	=	83
				F(14, 68)	=	169.95
Model	2556324.68	14	182594.62	Prob > F	=	0.0000
Residual	73058.8041	68	1074.39418	R-squared	=	0.9722
				Adj R-squared	=	0.9665
Total	2629383.49	82	32065.6523	Root MSE	=	32.778

puntaje	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
pearson	24.33014	35.05709	0.69	0.490	-45.62519	94.28547
toefl	74.33876	8.500742	8.74	0.000	57.37579	91.30172
toeic	609.1255	14.02444	43.43	0.000	581.1402	637.1109
masculino	30.46849	8.346725	3.65	0.001	13.81286	47.12412
partsub	-17.92809	8.289022	-2.16	0.034	-34.46857	-1.387602
media	18.70132	11.22829	1.67	0.100	-3.704391	41.10703
clasedos	62.65146	21.01486	2.98	0.004	20.71696	104.586
clasetres	41.58869	20.50822	2.03	0.046	.6651847	82.5122
clasecuatro	48.34822	21.06267	2.30	0.025	6.318304	90.37813
clasecinco	37.8067	21.91329	1.73	0.089	-5.920584	81.53398
claseseis	30.32068	40.59264	0.75	0.458	-50.68067	111.322
mmedia	20.44067	13.82929	1.48	0.144	-7.155237	48.03658
msuperior	29.46159	13.75562	2.14	0.036	2.012684	56.9105
Peñalolén	28.50172	34.27256	0.83	0.409	-39.8881	96.89154
_cons	-61.96268	24.33987	-2.55	0.013	-110.5321	-13.39323

Se ve que hay algunas variables no significativas, por lo que se hizo una selección de modelos final vía stepwise. Luego de esta selección de modelos, el modelo quedó igual.

Algunas observaciones y conclusiones:

- El r2 ajustado subió de 80% (modelo inicial) a 96% (modelo final).
- A los hombres les va mejor que a las mujeres en las mencionadas pruebas: contrario a lo esperado (que era igualdad estadística).
- La edad no influye en el puntaje: contrario a lo esperado (que influya positivamente).
- A los alumnos que estudiaron más en colegios particulares subvencionados les va peor que a los que lo hicieron en públicos y a los alumnos de colegios particulares no les va mejor que a los de públicos, que es contrario a lo esperado.
- A los alumnos que tenían nivel máximo educativo “enseñanza media” al dar el examen les va mejor que a los de educación básica: de acuerdo a lo esperado, pero a los alumnos de

otros niveles académicos no tienen diferencia con los de educación básica, lo que es contrario a lo esperado.

- Haber tenido al menos una asignatura dictada 100% en inglés no afecta el puntaje: contrario a lo esperado.
- No se encontró relación entre nivel socioeconómico (aproximado por ingreso en el hogar por persona en la época que el encuestado aprendió más inglés) y el puntaje: contrario a lo esperado.
- Haber vivido en países anglosajones no afecta el puntaje: contrario a lo esperado.
- La preparación específica para el examen no afecta el puntaje: contrario a lo esperado.
- Ser autodidacta no afecta el puntaje: contrario a lo esperado.
- El nivel educativo de la madre influye en el puntaje sólo si llegó hasta educación superior o media. Influye más de manera positiva si es educación superior: parcialmente de acuerdo a lo esperado (que mientras mayor fuese el nivel educativo, mayor puntaje).
- El nivel educativo del padre no afecta el puntaje: contrario a lo esperado.
- Las personas que le hablaban en inglés ya sea cuando aprendió más inglés o en su infancia/adolescencia no afecta el puntaje: contrario a lo esperado.
- Sólo haber residido en la comuna de Peñalolén en la época que aprendió más inglés afecta de manera positiva el puntaje (con respecto a la base Ancud), las demás no lo afectan. Contrario a lo esperado, ya que Peñalolén es más pobre (en general) que Ancud.

Con respecto al estudio de 2010 en Valdivia, acá se ve que la motivación (aproximada por la variable autodidacta) no influye en el nivel de inglés, y al igual que creían las encuestadas de ese estudio, el tener asignaturas 100% dictadas en inglés no influye en el nivel de inglés.

A diferencia del estudio mencionado de 2014 en Chillán, acá la ubicación geográfica en la mayoría de los casos no influye en el nivel de inglés (aunque en el presente trabajo se considera todo el mundo y en el estudio de 2014 sólo Chillán).

A diferencia del estudio hecho en México, acá haber estudiado en colegio particular no resulta en mejor desempeño.

A diferencia del estudio hecho en Perú, la motivación (acá aproximada por la variable autodidacta) no influye en el nivel de inglés.

A diferencia del estudio de EF, acá no hay diferencia geográfica en el puntaje (en general) ni é área.

Las conclusiones finales, que responden a la pregunta principal de la investigación (¿Cuáles son los factores que influyen en el nivel de inglés de los chilenos?), las respuestas, son:

- Género (hombres tienen mayor nivel).
- Haber estudiado en colegio particular subvencionado (de manera negativa).
- Haber llegado hasta enseñanza media completa.
- Que su madre haya llegado hasta enseñanza media completa o universitaria completa.
- Residir en Peñalolén cuando se aprende inglés.

## Literatura citada

- Ceballos, K; Dávila N; Espinoza J; Ramírez M (2014); “Factores que inciden en el aprendizaje del idioma inglés en los alumnos de segundo año medio de la ciudad de Chillán” (Tesis de pregrado, Universidad del Bío-bío).
- Astudillo, T (2010); “Factores que influyen en el aprendizaje del idioma inglés en alumnas del Liceo Santa María la Blanca de Valdivia” (Tesis de pregrado, Universidad Austral de Chile)
- Valenzuela M; Romero K; Vidal-Silva C; Philominraj A (2016); “Factores que influyen en el aprendizaje del idioma inglés de nivel Inicial en una universidad chilena”.
- Education First (2020), <https://www.ef.com/cl/epi/>
- Bastidas J; Muñoz G (2020); “Factores que influyen en el aprendizaje del inglés de los bachilleres de Pasto”.
- Santana J; García-Santillán A; Escalera-Chavez Milka (2016); “Variables que influyen sobre el aprendizaje del inglés como segunda lengua”.
- Gonzales F; Gonzales A; Trujillo N; Fernandez C (2020); “Factores influyentes en el aprendizaje del inglés en instituciones públicas”.
- Wooldridge J (2010): “Introducción a la econometría, un enfoque moderno”.
- StataCorp (2021). *Stata 17 Base Reference Manual*. College Station, TX: Stata Press.