



TÍTULO

ACCESO Y USO DE TECNOLOGÍAS DE LA INFORMACIÓN Y
COMUNICACIÓN EN ESPAÑA
ANÁLISIS E IDENTIFICACIÓN DE FACTORES DETERMINANTES

AUTOR

Osmar Capote Vázquez

Tutora	Esta edición electrónica ha sido realizada en 2022
Instituciones	Dra. D ^a . Mónica Carmona Arango
Curso	Universidad Internacional de Andalucía ; Universidad de Huelva
©	<i>Máster en Economía, Finanzas y Computación (2020/21)</i>
©	Osmar Capote Vázquez
Fecha documento	De esta edición: Universidad Internacional de Andalucía
	2021



**Atribución-NoComercial-SinDerivadas
4.0 Internacional (CC BY-NC-ND 4.0)**

Para más información:

<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>

ACCESO Y USO DE TECNOLOGÍAS DE LA INFORMACIÓN Y COMUNICACIÓN EN ESPAÑA: ANÁLISIS E IDENTIFICACIÓN DE FACTORES DETERMINANTES

by

Osmar Capote Vázquez

A thesis submitted in conformity with the requirements
for the MSc in Economics, Finance and Computer Science
University of Huelva & International University of Andalusia

uhu.es

un
i Universidad
Internacional
de Andalucía
A

Noviembre 2021

ACCESO Y USO DE TECNOLOGÍAS DE LA INFORMACIÓN Y COMUNICACIÓN EN ESPAÑA: ANÁLISIS E IDENTIFICACIÓN DE FACTORES DETERMINANTES

Osmar Capote Vázquez

Máster en Economía, Finanzas y Computación

Supervisora

Mónica Carmona Arango

Universidad de Huelva y Universidad Internacional de Andalucía

2021

Abstract

This paper examines socioeconomic and demographic factors, as well as technological equipment, as determinants of ICT access and use in Spanish households and influencing the so-called digital divide. For this purpose, microdata from the Survey on Equipment and Use of Information and Communication Technologies in Households (ETIC_H), prepared by the National Statistics Institute (INE) in Spain, for the year 2020, with a sample of 15343 people, are used. In order to advance the research, a Logit model is used, which belongs to the group of binary discrete choice models and is estimated by means of the maximum likelihood technique. The results obtained show some conclusions drawn from the estimation of the models, taking into account the hypotheses put forward in the research.

JEL classification: I25, I28, J24, J28, O52.

Key words: Information and Communication Technologies, Logit model, digital divide.

Resumen

Este trabajo examina los factores tanto socioeconómicos, demográficos, así como el equipamiento tecnológico como determinantes del acceso y uso de las TIC en los hogares españoles e influyentes en la denominada brecha digital. Para ello se utilizan los microdatos de la Encuesta sobre Equipamiento y Uso de Tecnologías de Información y Comunicación en los Hogares (ETIC_H), que elabora el Instituto Nacional de Estadística (INE) en España, para el año 2020, con una muestra de 15343 personas. Para el avance de la investigación se utiliza un modelo Logit el cual pertenece al grupo de modelos de elección discreta binario y se estima mediante la técnica de máxima verosimilitud. En los resultados obtenidos se pone de manifiesto algunas conclusiones que arrojan la estimación de los modelos teniendo en cuenta las hipótesis planteadas en la investigación.

Palabras claves: Tecnologías de la información y la comunicación (TIC), modelo Logit, brecha digital.

Agradecimientos

A mis padres que siempre han sido mi motor impulsor. Y en especial a mí pequeña Valeria y a mi eterna novia que tantas largas noches sin dormir han estado a mi lado para lograr este tan anhelado objetivo que es hacerme Máster.

Índice general

1.- Introducción.....	p.1
2.- Revisión de la literatura.....	p.6
2.1.- Conclusiones del estudio de la bibliografía.....	p.9
3.- Metodología.....	p.9
3.1.- Base de datos.....	p.9
3.2.- Variables.....	p.12
3.3.- Datos.....	p.16
3.4.- Modelo.....	p.24
3.5.- Modelo de elección discreta binaria.....	p.24
4.- Resultados.....	p.27
4.1. - Modelo logit.....	p.28
4.2.- Efectos marginales.....	p.31
5.- Conclusiones.....	p.35
6.- Referencias bibliográficas.....	p.36
7.- Anexos.....	p.38

Índice de tablas

Tabla 1. Clasificación por el INE de los apartados de la ETIC_H.	11
Tabla 2. Estructura de la encuesta por bloques.	11
Tabla 3 Variables demográficas.	13
Tabla 4. Variables socioeconómicas.	14
Tabla 5. Variables referentes al equipamiento/acceso a Internet con que dispone la vivienda. ...	15
Tabla 6. Descripción de las variables del <i>Modelo 1</i>	26
Tabla 7. Descripción de las variables del <i>Modelo 2</i>	26
Tabla 8. Descripción de las variables del <i>Modelo 3</i>	26
Tabla 9. Descripción de las variables del <i>Modelo 4</i>	26
Tabla 10. Descripción de las variables del <i>Modelo 5</i>	27
Tabla 11. Descripción de las variables del <i>Modelo 6</i>	27

Índice de figuras

Figura 1. Agrupamiento del número de personas encuestadas en la ETIC_H teniendo en cuenta a comunidades y ciudades autónomas y capitales de provincia en España.....	17
Figura 2. Estadísticos descriptivos de la variable dependiente “USO_INT”.....	19
Figura 3. Estadísticos descriptivos de las variables independientes identificadas a partir de las hipótesis planteadas en la investigación.	19
Figura 4. Relación entre variables identificadas (nivel de estudios y servicios usados de Internet) a partir de la <i>H1</i>	19
Figura 5. Relación entre variables identificadas (grupo de edad y servicios usados de Internet) a partir de la <i>H2</i>	20
Figura 6. Relación entre variables identificadas (ingreso neto del hogar y servicios usados de Internet) a partir de la <i>H3</i>	20
Figura 7. Relación entre variables identificadas (sexo y servicios usados de Internet) a partir de la <i>H4</i>	20
Figura 8. Correlación múltiple entre la variable dependiente (uso del Internet) e independiente (nivel de educación alcanzado) asociadas a la <i>H1</i>	22
Figura 9. Correlación múltiple entre la variable dependiente (uso del Internet) e independiente (grupos de edad) asociadas a la <i>H2</i>	22
Figura 10. Correlación múltiple entre la variable dependiente (uso del Internet) e independiente (ingresos netos del hogar) asociadas a la <i>H3</i>	23
Figura 11. Correlación múltiple entre la variable dependiente (uso del Internet) e independiente (sexo) asociadas a la <i>H4</i>	23
Figura 12. Correlación múltiple entre la variable dependiente (uso del Internet) e independientes (formas de conexión a internet: banda ancha y móvil respectivamente) asociadas a la <i>H5</i>	23
Figura 13. Correlación múltiple entre la variable dependiente (uso del Internet) e independientes (situación laboral, ocupación principal y actividad del establecimiento) asociadas a la <i>H6</i>	23
Figura 14. Estimación del modelo logit para el <i>Modelo 1</i> , donde interviene la variable independiente (nivel de estudios terminados) con la dependiente (uso del Internet).	28
Figura 15. Estimación del modelo logit para el <i>Modelo 2</i> , donde interviene la variable independiente (grupos de edades) con la dependiente (uso del Internet).	29

Figura 16. Estimación del modelo logit para el <i>Modelo 3</i> , donde interviene la variable independiente (ingresos netos del hogar) con la dependiente (uso del Internet).....	29
Figura 17. Estimación del modelo logit para el <i>Modelo 4</i> , donde interviene la variable independiente (sexo) con la dependiente (uso del Internet).....	30
Figura 18. Estimación del modelo logit para el <i>Modelo 5</i> , donde interviene la variable independiente (tipo de conexión fija y tipo de conexión móvil a Internet) con la dependiente (uso del Internet).....	30
Figura 19. Estimación del modelo logit para el <i>Modelo 6</i> , donde interviene la variable independiente (tipo de conexión fija y tipo de conexión móvil a Internet) con la dependiente (uso del Internet).....	31
Figura 20. Efecto marginal luego de estimar el <i>Modelo 1</i>	32
Figura 21. Efecto marginal luego de estimar el <i>Modelo 2</i>	32
Figura 22. Efecto marginal luego de estimar el <i>Modelo 3</i>	33
Figura 23. Efecto marginal luego de estimar el <i>Modelo 4</i>	33
Figura 24. Efecto marginal luego de estimar el <i>Modelo 5</i>	34
Figura 25. Efecto marginal luego de estimar el <i>Modelo 6</i>	35

1. Introducción

Las Tecnologías de la Información y las Comunicaciones (TIC) han evolucionado a lo largo de los últimos años. Estas ofrecen diversas ventajas entre las que se destacan tener un mayor acceso a la información, además una reducción de los costes en el sector laboral, así como la existencia de una mayor conectividad entre las personas. Pese a lo anterior, existen diferencias en el acceso y conocimiento de uso de las TIC lo cual se conoce como “brecha digital”. Este fenómeno, denominado brecha digital, no es otra cosa que las desigualdades que existen entre las personas, comunidades o países que tienen acceso y uso de las TIC. Y, sobre todo, de Internet como una parte rutinaria de su vida diaria y aquéllas que no tienen acceso a las mismas y que, aunque lo tengan, no saben utilizarlas de manera óptima para su beneficio.

El *World Resources Institute* (2000) apoyado por varias de las más importantes empresas internacionales informáticas explica así esa hendidura que lejos de atenuar, está profundizando las desigualdades en el planeta. Se señala la existencia en todos los países de porcentaje de personas que tienen acceso a la mejor tecnología y a las mejores redes de información y comunicación, mientras que existe otro grupo que no tienen acceso a estas tecnologías, por aspectos derivados de múltiples factores (sociales, culturales, raciales, educativos, económicos, geográficos, laborales, de género y de edad, entre otros), y la diferencia entre estos dos grupos es la nombrada brecha digital (Mendoza-Ruano & Caldera-Serrano, 2014).

Partiendo de que desde el propio tema de la investigación hace alusión de ello, sería fructífero y provechoso aclarar que al estudiar el rol de las TIC es pertinente diferenciar entre acceso y uso. Mientras que el acceso se vincula a la disponibilidad de las tecnologías/equipamiento informático, el uso tiene que ver en cómo y con qué fines se utilizan. Tanto las TIC, el acceso y uso de las mismas, así como la tan popular y definida brecha digital, no se pueden analizar como aspectos aislados. Los mismos son paradigmas relevantes que influyen directamente en la muy nombrada sociedad de la información. Es así que, es factible definir dos brechas digitales, una denominada “primera” referida al acceso a las TIC y otra llamada “segunda” relacionada a las desigualdades en la capacidad de obtener beneficios de las mismas (Alderete & Formichella, 2020).

El uso de las TIC en los hogares ha crecido en los últimos años, si bien sigue existiendo una brecha entre los usuarios y no usuarios (brecha digital) de estas tecnologías que se puede atribuir a una serie de factores: la falta de infraestructura (en particular en las zonas rurales), la falta de conocimientos de informática y habilidades necesarias para participar en la sociedad de la información, o la falta de interés en lo que la sociedad de la información puede ofrecer. A pesar de los elementos determinantes anteriormente expresados, es indiscutible e imparable la proliferación que están teniendo las TIC en los hogares. A tal punto que la sociedad cada vez se vuelve más dependiente de los avances tecnológicos.

Cada ciudadano hoy en día debe estar capacitado para vivir y trabajar en la sociedad de la información. Como España no está aislada de todo el acontecer tecnológico que sucede en el mundo, se ha dado a la tarea de potenciar su informatización. Llegar a alcanzar que los hogares españoles vean en las TIC una oportunidad de superación, facilidades comunicativas tanto en lo profesional como en términos personales. Que puedan lograr un buen aprovechamiento de las mencionadas, que la sociedad cree una cultura digital donde las TIC contribuyan a un continuo desarrollo sostenible de la nación.

Por lo que, para lograr el progreso de una sociedad digital, un grupo de entidades, organismos estatales, organizaciones nacionales y multinacionales españolas promueven la creación de agendas digitales. También se llevan a cabo proyectos mediante encuestas, entrevistas personales a nivel nacional que arrojan como resultado base de datos con información estadística de gran importancia. Esos resultados funcionan como base para contrastar tanto para en el interior de España, para la Unión Europea (UE), así como a nivel internacional. Dichos proyectos se implementan cada año con el objetivo de obtener datos en lo referente al desarrollo y evolución del equipamiento, acceso y uso de Internet en los hogares. La formación propiciada por las instituciones designadas, así como las agendas digitales, *webinars*, *masterclass* y proyectos que se lleven a cabo en España para ganar en aras de hacer un buen uso de las TIC tienen un mismo objetivo que es especialmente la realización de actividades que apoyen la alfabetización digital y la accesibilidad (INE, 2020).

Al hacer un análisis en lo que respecta la brecha digital en los hogares españoles se identifican un grupo de factores determinantes e impulsores de la misma. Ya sea el equipamiento en los hogares, así como el uso del ordenador, Internet, comercio electrónico siendo aspectos que

condicionan en gran medida la vida diaria de las personas, en el lugar de trabajo, en los centros de educación y en los hogares. Analizar el uso de productos TIC por las personas, en particular el uso de Internet según la edad y el sexo nos proporciona información del perfil de la persona que los utiliza, de las diferencias de género en el uso, y del grado de desarrollo de las TIC en la sociedad. El estudio de lo anterior permite contrastar que existen notables diferencias que son cada vez más numerosas respecto al acceso y uso de las TIC.

Dentro de las instituciones españolas dedicadas a fomentar, estudiar, analizar, publicar y brindar información referente al equipamiento y uso de las TIC en los hogares se encuentra el Instituto Nacional de Estadística (INE). Instituto que gracias a su constante trabajo brinda anualmente gran volumen de información detallada mediante la Encuesta sobre Equipamiento y Uso de Tecnologías de Información y Comunicación en los Hogares (ETIC_H). Se han logrado obtener datos, conclusiones estadísticas como, por ejemplo: en el año 2020 en España, el 91,3% de la población de 16 a 74 años ha utilizado Internet de manera frecuente (al menos una vez por semana en los últimos tres meses). Es ligeramente superior el porcentaje de mujeres (91,4%) como usuarios frecuentes de Internet al de hombres (91,1%).

Los porcentajes más altos de usuarios frecuentes de Internet (tanto en hombres como en mujeres) corresponden a las personas más jóvenes, un 99,9% de los hombres y un 99,6% de las mujeres de 16 a 24 años son usuarios frecuentes de Internet. Al aumentar la edad descende el uso de Internet de manera frecuente tanto en hombres como en mujeres. El porcentaje más bajo corresponde al grupo de edad de 65 a 74 años. Un 64,0% de los hombres y un 64,4% de las mujeres de 65 a 74 años utilizan Internet de manera frecuente.

A medida que se eleva el nivel de estudios aumenta el porcentaje de usuarios frecuentes de Internet tanto en hombres como en mujeres. Como, por ejemplo: en el nivel de estudios de doctorado universitario un 99,1% de hombres (de 16 a 74 años) son usuarios frecuentes de Internet y un 100% de mujeres de este mismo nivel de estudios. El porcentaje más bajo corresponde al nivel de estudios de analfabetos y primaria incompleta (INE, 2020).

Además, se plantea que el 92,5% de los niños de entre 10 y 15 años se conectan desde su casa y el 70,9% desde el centro de estudios; aunque de manera general, se conoce que los adolescentes de 12 a 17 años tienen acceso a las tecnologías, pero no se confirman los resultados de conexión

excesiva. Asimismo, se evidencia una menor utilización de tabletas y consolas, que habría que relacionarla en estudios posteriores con el nivel de acceso que tienen a las mismas. La mayor parte del tiempo dedicado a las actividades con las TIC se da en el hogar familiar, seguido del colegio y la calle (Gairín Sallán & Mercader, 2018).

Se plantea que los adolescentes españoles pasan una media de entre dos y tres horas diarias utilizando algún tipo de dispositivo tecnológico, siendo cada vez más los que disponen de teléfono móvil propio (69.8%) y el uso del ordenador e Internet lo que representa el 94.9% y 95.2%, respectivamente. También, se evidencia que el 87.1% de los jóvenes dispone de un ordenador y el 84.2% de móvil, utilizando ambos dispositivos tanto en el contexto académico como social. Los jóvenes los utilizan mayoritariamente en casa (52.36%), la calle (16.25%) y el instituto (13.49%) (Díaz-Vicario, Mercader Juan, & Gairín Sallán, 2019).

Otro estudio señala que la falta de acceso y el analfabetismo digital de algunas familias y docentes son factores que influyen negativamente en el uso de las TIC. Siendo lo anterior consecuencia de la falta de formación en el uso de las TIC por parte de los maestros. También, se describe sobre el uso de las TIC para la comunicación entre padres y maestros, así como para mejorar la participación familiar en las instituciones educativas; teniendo como inconvenientes, la existencia de desigualdades estructurales (donde por problemas económicos o sociales no se tiene acceso a las tecnologías) o por desigualdades de agencia (cuando se tiene acceso, pero no se sabe qué hacer con ellas). Lo anterior es demostrado mediante entrevistas que arrojan como resultado que el 36% de los entrevistados destacan la falta de equipamiento y/o conexión que sufren algunas familias en la mayoría de los centros (Bordalba & Bochaca, 2018).

Es válido destacar que, existen evidencias de que los estudiantes que no disponen de ordenador/tabletas para uso personal en casa, presentan menores porcentajes sobre el uso diario de las TIC en el aula. A medida que aumenta la edad del estudiante y el nivel de estudios, disminuye el porcentaje del impacto del uso de las TIC en el aula sobre el aprendizaje y se incrementa a medida que se refieren a actividades más complejas. De manera general, los estudiantes refieren tener un mayor conocimiento y competencia digital que el profesorado (Moreira, Romero, & García, 2018).

Por lo que, atendiendo a lo anteriormente expresado, se plantea como **objetivo principal de la presente investigación**: *examinar los determinantes del acceso y uso de las TIC en los hogares, es decir, que factores llevan a los españoles a hacer uso o no de las TIC. Y una vez tomada la decisión de usar estas tecnologías, cuáles son los factores que hacen que los hogares decidan usar más o menos en esta partida.*

Para dar cumplimiento al objetivo principal se desglosaron los siguientes objetivos específicos:

1. Profundizar en el estudio del problema del acceso y uso de las TIC a través de la investigación bibliográfica.
2. Manipular los microdatos facilitados por el INE atendiendo a sus variables.
3. Examinar la relación que existe entre las variables demográficas (sexo y edad), socioeconómicas (ingresos netos mensuales, estudios terminados, situación laboral y ocupación principal), equipamiento para acceso a Internet (conexión por banda ancha o móvil) y el uso del Internet que permita medir la difusión de las TIC en España.
4. Elaborar un modelo econométrico que permita describir las relaciones entre las variables demográficas, socioeconómicas, equipamiento para acceso a Internet y el uso del Internet que afecta el grado de difusión de las TIC en España.
5. Analizar los resultados arrojados por el modelo estimado teniendo en cuenta las hipótesis a constatar.
6. Derivar conclusiones sobre la magnitud del problema de estudio.

Para poder llevar a cabo este análisis se utiliza un conjunto de datos que se han elaborado partir de los microdatos que provee el INE en su ETIC_H, para el año 2020 y serán explicados en epígrafes posteriores de este documento. Por su parte, la ETIC_H aporta información muy detallada acerca del acceso y uso de las TIC en los hogares españoles habitados por al menos una persona de entre 16 a 74 años de edad. Además de los ingresos mensuales netos del hogar, las características sociodemográficas, económicas, la formación, nacionalidad y la información que tiene que ver con el equipamiento con el que disponen los diferentes miembros que conforman el hogar. Es válido destacar que, para el análisis sobre el acceso y uso de las TIC, en la encuesta se

valoran como grupos de edades de 16 a 24 años, de 25 a 34 años, de 35 a 44 años, de 45 a 54 años, de 55 a 64 años, de 65 a 74 años y de 75 o más años (EL-DERECHO, 2020). Partiendo de esta información, así como de los trabajos científicos estudiados que se expondrán en el epígrafe 2 se plantean las siguientes hipótesis a demostrar para el logro del objetivo principal:

H1. El nivel de educación alcanzado por los individuos en España influye positivamente en el uso del Internet.

H2. Las personas jóvenes hacen mayor uso del Internet en comparación con las demás.

H3. En España, el ingreso neto del hogar influye positivamente en el uso del Internet en los hogares.

H4. En los hogares españoles los hombres hacen mayor uso del Internet que las mujeres.

H5. Existe mayor impacto en el uso del Internet mediante la conexión fija a Internet que mediante la conexión móvil en los hogares españoles.

H6. El estatus socioeconómico (situación laboral, ocupación principal y actividad del establecimiento) se relaciona positivamente con el uso del internet.

Para ello, este documento se estructura de la siguiente manera. Una vez introducido el tema de análisis y los objetivos, en el segundo apartado se realiza una revisión de los principales trabajos publicados acerca de las decisiones del acceso y uso del Internet. En la tercera parte, se explican los datos que se han utilizado para este análisis, así como las distintas variables con las que se van a trabajar, además se explica el modelo a utilizar que facilitará contrastar las hipótesis planteadas. Con todo esto se pasará a la cuarta parte, donde se extraen los resultados del mismo. Por último, en la quinta y última parte se extraen las conclusiones que se derivan de la investigación.

2. Revisión de la literatura

Los siguientes antecedentes constituyen investigaciones que presentan importantes datos y conclusiones sobre el tema de investigación planteado.

Millán realiza un estudio sobre el grado de utilización de las TIC por parte de diferentes tipos de autónomos y cómo afecta a sus ingresos. Para el avance de la misma hace uso de una encuesta española donde se recogen datos de 35 países europeos. Para contrastar los supuestos planteados utiliza el modelo Heckman donde obtiene como resultado que los ingresos aumentan con el nivel de uso de las TIC, pero solo a partir de un umbral de utilización que suponga al menos el 25% del tiempo (Millán, Lyalkov, Burke, Millán, & van Stel, 2019).

Domínguez en su investigación tiene como objetivo identificar los factores que inciden en la disponibilidad y el uso del celular inteligente en el sector rural de México. Entre las variables utilizadas para contrastar los supuestos planteados se encuentran la edad, el ingreso, género y nivel de escolaridad. Los resultados del modelo Heckman, el cual fue el utilizado, indican que estas variables son factores que determinan el acceso y uso de un celular inteligente. Además, la disponibilidad de conexión fija a Internet tiene un impacto negativo y significativo en el uso de aplicaciones relacionadas con mensajes, juegos y redes sociales, lo que sugiere que el celular inteligente es un sustituto para estas actividades (Domínguez & Navarro, 2019).

Zamora incide en el estudio e identificación de los factores que determinan el uso de Internet por parte de las personas con discapacidad, así como la existencia de brecha digital. Para alcanzar los resultados obtenidos utiliza el modelo Heckman. Se realizó una estimación por el método *probit*, de modo que se obtenga la probabilidad de usar o no las TIC y la existencia de brechas digitales obteniéndose como resultado que se cumplen las hipótesis planteadas (Zamora Salas & Rios Espinoza, 2019).

Catón y otros, realizan una revisión sistemática que examina el concepto “brecha digital” en la literatura científica, con el objetivo de conocer los factores que influyen en la misma mediante un consenso de diferentes autores con relación al desarrollo y la tecnología, incluyendo el papel que juega a nivel internacional la educación. En este estudio bibliográfico se utilizaron técnicas de búsqueda de información en la literatura. Entre los factores analizados se encuentran los ingresos, la educación, edad y el sexo, además del capital humano, así como el contexto familiar y social. El análisis de varios autores arroja como resultado que los factores que principalmente aparecen asociados a la brecha digital son de tipo socioeconómico como lo son el género, la edad y el uso de video-juegos (Catón Mayo, Cañón Rodríguez, & Grande de Prado, 2016).

Souza y otros autores describen lo que representa la brecha digital en las pequeñas y medianas empresas de Brasil mediante un indicador del uso de las TIC. Para el avance de la investigación utilizan una encuesta sobre las TIC en las empresas publicada en el 2015. Mediante ecuaciones estructurales, se creó un índice para medir la intensidad del uso de las TIC en las organizaciones y se verificaron los factores que lo afectan mediante regresiones lineales. El principal resultado arrojado permite interpretar que el acceso a la tecnología e Internet, así como las competencias y actitudes en materia de TIC son factores correlacionados con un uso más intenso de las TIC por parte de las organizaciones (Souza, Siqueira, & Reinhard, 2017).

Cuervo y Menéndez realizan en su investigación un análisis de los factores determinantes del uso de las TIC en España. Para el avance de la investigación se realiza una encuesta telefónica a personas mayores de 14 años en España. El análisis se desarrolla con modelos Logit, teniendo como variables dependientes las relacionadas al uso de Internet y la telefonía móvil. Los resultados del trabajo confirman que el mayor uso del Internet está asociado a los más jóvenes y que los hombres tienen mayor probabilidad de uso del Internet que las mujeres (Cuervo & Menéndez, 2005).

En el trabajo realizado por Jiménez se plantea como objetivo analizar las TIC como factores que inciden en un aumento del uso del Internet en Oaxaca, México. Para el avance de la investigación se tomaron los datos de la Encuesta Nacional sobre Disponibilidad y Uso de las TIC en los Hogares (Endutih) realizada por el Instituto Nacional de Estadística y Geografía (Inegi). Para alcanzar los resultados de la misma se utilizó el modelo de regresión lineal múltiple con MCO. Esto permitió conocer que los mensajes instantáneos, los tutoriales y la música son los que tienen un mayor impacto en el aumento del uso del Internet (Jiménez García & Pérez Soto, 2020).

Un estudio realizado por Rodríguez tiene como objetivo explicar la brecha digital en México desde una perspectiva de la dotación de capital TIC, del hogar o del individuo. Para el desarrollo de la investigación se hace uso de un modelo Logit con base en microdatos de una encuesta nacional de telecomunicaciones, para el año 2016. Como resultado se obtiene que la dotación del capital de telecomunicaciones es crucial, además del nivel educativo, grupo de edad, tamaño del hogar, ingreso económico y urbanización (Rodríguez Ruiz, 2019).

2.1 Conclusiones del estudio de la bibliografía

Sobre la base de los estudios realizados anteriormente acerca del acceso y uso de las TIC se evidencia que existe similitud en cuanto a los factores a evaluar. Entre las variables que predominan se encuentran las demográficas (sexo y edad), socioeconómicas (ingresos, estudios terminados, situación laboral y ocupación), equipamiento para el acceso a Internet (formas de conexión a Internet) y el uso del Internet, las cuales influyen en la brecha digital. Se utilizaron encuestas nacionales e internacionales para recoger los datos y analizarlos. Entre los modelos econométricos utilizados se encuentra el modelo de regresión lineal, modelo Heckman y modelo Logit que permitieron el análisis de los resultados para contrastar los supuestos planteados en cada contribución.

Con referencia a lo anterior, resulta interesante la utilización de un modelo Logit teniendo en cuenta la naturaleza de la variable dependiente de las hipótesis del presente trabajo, así como el principal objetivo trazado. Para el avance de la investigación se utiliza la ETIC_H que llevó a cabo el INE en España en el 2020, que engloban información referente al acceso y uso de las TIC en los hogares incidiendo en las variables antes mencionadas.

3. Metodología

3.1. Base de datos

Como ya se expresó anteriormente, la base de datos que se va a utilizar en la presente investigación es la ETIC_H que implementa el INE en España. La encuesta, con base 2006-2020, es un estudio que se realiza cada año y que obtiene su información a través de entrevistas personales a una muestra teórica inicial de 25.837 hogares españoles aproximadamente de los cuales 21.890 aproximadamente, el 84,7% resultaron encuestables. De las 21.890 viviendas encuestables de partida, 15.343 viviendas cumplieron un cuestionario completo (Vivienda encuestada o muestra neta). Por tanto, la tasa de falta de respuesta se sitúa en un 29,9% y sobre todo señalar que son justamente esas 15.343 viviendas/personas (1 persona encuestada por vivienda) analizadas como observaciones en las cuales se va a fundamentar el presente trabajo para contrastar los supuestos planteados.

La ETIC_H realizada por el INE en colaboración con el Instituto de Estadística de Cataluña (IDESCAT) y el Instituto de Estadística y Cartografía de Andalucía (IECA) sigue las recomendaciones metodológicas de la oficina de estadística de la Unión Europea (Eurostat). Es la única fuente en su género cuyos datos son estrictamente comparables no sólo entre los países miembros de la Unión, sino además en otros ámbitos internacionales (INE, 2020).

Al objeto de lograr una mayor comparabilidad con los datos que publica Eurostat, los resultados que se presentan a partir del 2006, se refieren a viviendas habitadas por al menos una persona de 16 a 74 años de edad. No obstante, en los ficheros de microdatos se ofrecen los resultados completos de la población de 16 y más años. Asimismo, en la publicación se incluyen tablas adicionales relativas tanto a menores (de 10 a 15 años) como de personas de 75 y más años.

Como objetivos principales, la ETIC_H busca (INE, 2020):

- Conocer el equipamiento en tecnologías de la información y la comunicación de los hogares españoles (conexión a Internet, telefonía fija y móvil, y equipamiento informático).
- Analizar el uso y los nuevos hábitos de la población española respecto a Internet: actividades realizadas, utilización del comercio electrónico, capacidades y conocimientos informáticos, medidas adoptadas sobre protección, seguridad y privacidad, relaciones con la administración electrónica y el uso del Internet de las cosas.
- Servir de base para establecer comparaciones entre España y otros países y satisfacer los requerimientos de información de organismos internacionales. Es la única fuente en su género cuyos datos son estrictamente comparables no sólo entre países de la Unión Europea, sino además en otros ámbitos internacionales.

En cuanto a su elaboración, el cuestionario utilizado está formado por 12 bloques que pueden agruparse en tres grandes apartados (ver tabla 1) que a su vez se componen en 64 preguntas referentes al acceso y uso de las TIC en los hogares españoles. Es la misma que realiza el organismo europeo de estadística (EUROSTAT) para así conseguir que todos los países miembros de la Unión Europea realicen el mismo análisis.

Tabla 1. Clasificación por el INE de los apartados de la ETIC_H.

No.	Clasificación
1	Información TIC sobre la vivienda
2	Información TIC sobre todos los menores de 10 a 15 años
3	Información TIC sobre la persona mayor de 16 años seleccionada en la vivienda

Fuente: elaboración propia a partir de la ETIC_H del INE.

Para el análisis de los datos la investigación se enfoca en los apartados 1 y 3, ya que se va a estudiar todo lo referente a información TIC sobre la vivienda, así como la información TIC sobre la persona de 16 años o más, seleccionada en la vivienda. El apartado 1 contempla los bloques I, II y III del cuestionario, mientras que el apartado 3 contempla los bloques del V al XII.

Dentro de los resultados de la ETIC_H, se encuentran los ficheros de microdatos los cuales están agrupados como era de esperar por años e incluyen en cada fichero su diseño de registro propio, así como en el caso de la ETIC_H en el año 2020 que contiene un fichero “cuestionario” y un fichero de los “miembros de hogar”. En el fichero de cuestionario, se incluyen tantos registros como hogares/personas (una persona encuestada por vivienda) hayan colaborado, identificándose con la variable “PERSONA”. Es por ello que atendiendo a todo lo anteriormente expresado la información e incluso la estructura de la propia encuesta se divide en los 12 bloques presentados en la tabla 2.

Tabla 2. Estructura de la encuesta por bloques.

No.	Descripción
1	Composición del hogar.
2	Equipamiento de la vivienda principal en productos de tecnologías de información y comunicación.
3	Acceso a internet de la vivienda
4	Uso de ordenador e internet por los niños (10 a 15 años)
5	Uso de teléfono móvil e internet por la persona seleccionada
6	Administración electrónica
7	Comercio electrónico
8	Conocimientos informáticos
9	Privacidad y protección de datos personales
10	Confianza, seguridad y privacidad en internet
11	Internet de las cosas

12	Características socioeconómicas de la persona seleccionada
----	--

Fuente: Elaboración propia a partir de la ETIC_H del INE.

En temas de difusión de los resultados mencionar que las tablas estadísticas de los resultados de la encuesta junto con el informe metodológico y otra información complementaria (ficheros de microdatos anonimizados, tablas de errores de muestreo, cuestionario, análisis de la falta de respuesta, metodología estandarizada, glosario de términos, nota de prensa, infografía estática e infografía dinámica o *Tableau*) se publican en el sitio web del INE¹. Señalar que el número de años comparables de las series temporales publicadas en 2020 es 15 (desde 2006, año en el que se convergió con la metodología europea) (INE, 2020).

Es válido resaltar, que el ámbito poblacional se incluye completamente en el fichero cuestionario descrito anteriormente. Es por ello que en este trabajo se va hacer uso del fichero de cuestionario. El INE proporciona la información en un fichero plano, por lo que previamente se ha tenido que normalizar los datos mediante la versión 16 del programa STATA. Dicho procedimiento contribuye a poder trabajar con los datos, así como extraer la información y por lo tanto conocimiento de ellos. El proceso llevado a cabo para su creación y la preparación de la base de datos para su utilización se presenta en el anexo 1.

3.2. Variables

En la presente investigación, se analizan las decisiones de acceso y uso de las TIC. Para ello se introducen las variables explicativas presentadas en la revisión de la literatura. Desde un punto de vista econométrico para realizar la estimación se elabora un modelo de elección discreta binario, específicamente un modelo Logit. El mismo permitirá identificar los factores determinantes que llevan a los españoles a hacer uso o no de las TIC. Las variables que se van a utilizar son las que se han mencionado en el epígrafe 2.1 de este trabajo, las cuales toman diferentes valores según la información que aporte a nuestro modelo econométrico.

Partiendo de las hipótesis planteadas al inicio de la investigación, se identifica como variable dependiente que incide en ellas: “USO_INT” la cual es una variable dicotómica que toma valores

¹ www.ine.es

cero o uno, no contiene “*missing values*” y es utilizada para representar si alguna vez se ha hecho uso del Internet en los hogares españoles.

Cabe destacar que, de manera general, algunas de las variables originales han sido modificadas en cuanto a su contenido, así como la creación de nuevas variables. A continuación, en la tabla 3 se muestran las variables demográficas utilizadas, en la tabla 4 las variables socioeconómicas y en la tabla 5 las variables referentes al equipamiento/acceso a Internet. Además, se detallan los cambios realizados en cada una de ellas, este procedimiento de modificación queda completamente expuesto en el anexo 2.

Tabla 3 Variables demográficas.

Variables explicativas	Nº valores	Valores válidos
SEXO Sexo de la persona seleccionada	2	1=H 6=M Variable original de la ETIC_H. Toma los valores modificados: 1=H 0=M
EDAD Edad de la persona seleccionada	-	Personas de 16 o más años Variable original de la ETIC_H
GROUP_EDAD Edad de las personas agrupadas por rango de edades	7	Personas con edad entre 16 y más años Variable creada a partir de datos de la variable EDAD 1= De 16 a 24 años 2= De 25 a 34 años 3= De 35 a 44 años 4= De 45 a 54 años 5= De 55 a 64 años 6= De 65 a 74 años 7= De 75 o más años Toma valores del 1-7

Fuente: Elaboración propia a partir de ETIC_H del INE.

Tabla 4. Variables socioeconómicas.

Variables explicativas	Nº valores	Valores válidos
ING_HOG Ingresos mensuales netos del hogar	6	1= Menos de 900 euros 2= De 900 a menos de 1.600 euros 3= De 1.600 a menos de 2.500 euros 4= De 2.500 a menos de 3.000 euros 5= 3.000 o más euros 6=NS/NR (No sabe/ No responde) Variable original de la ETIC_H. Toma valores de 1-6
NIVELEST Estudios terminados, código	10	0= Analfabetos y estudios primarios incompletos 1= Educación Primaria 2= Primera etapa de la Educación Secundaria y similar (Bachiller elemental, EGB, ESO...) 3= Segunda etapa de la Educación Secundaria y similar (Bachiller Superior, BUP, Bachillerato, FPI, FP de Grado Medio...) 4= Educación postsecundaria no superior (CdPN) 5= Formación Profesional de Grado Superior (FP II) y títulos propios de universidades de duración igual o superior a 2 años. 6= Grados universitarios de 240 créditos ECTS (Bolonia), diplomados universitarios, títulos propios universitarios de experto o especialista y similares. 7= Grados universitarios de más de 240 créditos ECTS (Bolonia), licenciados, másteres y especialidades en Ciencias de la Salud por el sistema de residencia y similares. 8= Título de Doctorado. 9= No se puede codificar. Variable original de la ETIC_H. Toma valores de 0-9
SIT_LAB Situación laboral y profesional en la que se encuentra	10	1=Trabajando por cuenta ajena con contrato indefinido 2=Trabajando por cuenta ajena con contrato temporal 3= Trabajando por cuenta propia 4= Parado 5= Estudiante 6=Jubilado o prejubilado 7=Incapacitado permanente 8=Realizando tareas de voluntariado social 9=Labores del hogar 0= otra situación

		Variable original de la ETIC_H. Toma valores de 0-9
ACTIV Actividad del establecimiento	12	1=Agricultura, silvicultura y pesca 2=Industrias extractivas, manufactureras y otras industrias 3=Construcción 4=Comercio al por mayor y al por menor, transporte y hostelería 5=Información y comunicaciones 6=Actividades financieras y de seguros 7=Actividades inmobiliarias 8=Servicios empresariales 9=Administración pública, defensa, educación, actividades sanitarias y de servicios sociales 10=Otros servicios. 11= No se puede codificar Blanco (SIT_LAB ≠1,2,3) Variable original de la ETIC_H. Toma valores de 1-11 o Blanco (SIT_LAB ≠1,2,3)
OCUPACION1 Ocupación principal: Manual / No manual	3	1=Manual 6=NO Manual Blanco (SIT_LAB ≠1,2,3) Variable original de la ETIC_H. Toma los valores modificados: 1= Manual 0= NO Manual
OCUPACION2 Ocupación principal: Trabajador TIC / No TIC	3	1=TIC 6=NO TIC Blanco (SIT_LAB ≠1,2,3) Variable original de la ETIC_H. Toma los valores modificados: 1= TIC 0= NO TIC

Fuente: Elaboración propia a partir de ETIC_H del INE.

Tabla 5. Variables referentes al equipamiento/acceso a Internet con que dispone la vivienda.

Variables explicativas	Nº valores	Valores válidos
INTEFOR1 Forma de conexión fija a Internet en la vivienda banda ancha= (ADSL, red	4	1=SI 6=NO 9=NS/NR Blanco VIV_INTER ≠1

de cable, fibra, WIFI público...)		Variable original de la ETIC_H. Toma los valores modificados: 1=SI 0=NO 2=NS/NR
INTEFOR2 Forma de conexión móvil a Internet en la vivienda banda ancha= (dispositivo mano móvil, vía modem USB o tarjeta)	4	1=SI 6=NO 9=NS/NR Blanco VIV_INTER ≠1 Variable original de la ETIC_H. Toma los valores modificados: 1=SI 0=NO 2=NS/NR

Fuente: Elaboración propia a partir de ETIC_H del INE.

3.3. Datos

El diseño muestral desarrollado para llevar a cabo la ETIC_H sigue un tipo de muestreo trietápico con estratificación de las unidades de primera etapa. Las unidades de primera etapa son las secciones censales. Las unidades de segunda etapa, las viviendas familiares principales y, en tercera etapa, se selecciona una persona de más de 15 años en cada vivienda. Así mismo, en cada vivienda se investiga a todos los menores de 10 a 15 años.

Para la selección de la muestra, se utilizó un marco de áreas formado a partir de la relación de secciones censales existentes con referencia noviembre de 2015. No obstante, las variaciones que se producen en el tiempo, se incorporan periódicamente a la muestra. En la selección de unidades de segunda etapa se ha utilizado la relación de viviendas familiares principales en cada una de las secciones seleccionadas para la muestra obtenidas de la explotación del Padrón Continuo de habitantes más actualizado disponible. Las secciones se agrupan en estratos, dentro de cada comunidad autónoma, de acuerdo con el tamaño del municipio al que pertenecen (INE, 2020).

Una vez se analiza el *core* de la puesta en práctica del diseño muestral y de señalar el tipo de muestreo usado, no se podía dejar de explicar la manera en que se realiza la selección de la muestra. De la misma cabe destacar que para llevar a cabo la encuesta, la selección de las unidades de primera etapa en cada estrato se realizó con probabilidad proporcional al tamaño de cada sección. En segunda etapa las viviendas se han seleccionado con igual probabilidad,

mediante un muestreo sistemático con arranque aleatorio, entre las viviendas de la sección, procedimiento que conduce a obtener muestras autoponderadas de viviendas en cada estrato. Por último, en una tercera etapa, y dentro de cada vivienda, se elige con probabilidad igual una persona entre las de 16 y más años.

Para dar respuesta a las hipótesis planteadas en la investigación se utilizaron datos de la ETIC_H, del año 2020. La cobertura geográfica de la encuesta fue nacional, estatal y segmentada de cierto modo por la cantidad de habitantes en los municipios (hábitat). Su objetivo principal es obtener datos comparativos sobre la disponibilidad en los hogares de equipamiento para comunicarse y obtener o enviar información, así como sobre el uso que los miembros del hogar hacen de dichas tecnologías. Para el estudio la ETIC_H está encuadrada en las estadísticas comunitarias (Comunidades Autónomas “CCAA”) sobre la sociedad de la información, por lo que está sometida a los reglamentos comunitarios (INE, 2020).

Como se puede observar en la figura 1 se agruparon en siete bloques/estratos asociados a diferentes colores para así poder de una manera más simple identificar el número de personas (de 16 a 74 años) encuestadas. Esta operación estadística está incluida en el Plan Estadístico Nacional (PEN) del 2017-2020 como soporte del sistema de planeación de desarrollo regional, lo cual se compone de varias entidades para integrar proyectos de gran alcance con efectos que traspasan los límites de dos o más entidades federativas.

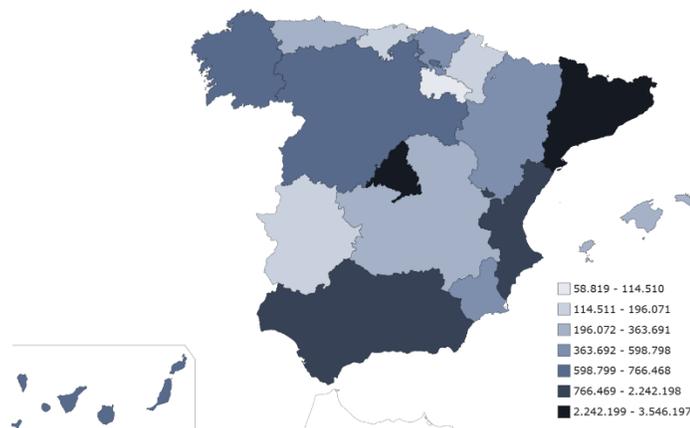


Figura 1. Agrupamiento del número de personas encuestadas en la ETIC_H teniendo en cuenta a comunidades y ciudades autónomas y capitales de provincia en España. Fuente: (INE, 2020).

En las estimaciones econométricas se incluyeron a los individuos de 16 y/o más años de edad, quienes forman parte del hogar en la ETIC_H, para el año 2020. Además, si en la vivienda conviven niños de 10 a 15 años se efectúa, para todos ellos, una serie mínima de preguntas relacionadas con este tema. En la investigación se tiene en cuenta el resumen y evolución de las principales variables así como información detallada sobre el equipamiento y el acceso a Internet en la vivienda y en lo concerniente a las personas; sobre uso de Internet, actividades desarrolladas en Internet, conocimientos y habilidades informáticas, comercio electrónico, relación con la administración electrónica, seguridad, privacidad, protección y confianza en Internet, así como el uso del Internet de las cosas (INE, 2020).

La utilización de las nuevas tecnologías por parte de los españoles se encuentra, en general, muy extendida. El 91% de los hogares españoles tenía acceso a Internet en el año 2019. En los hogares con hijos dependientes los porcentajes son más elevados, con un 99% en todos los tipos de hogares. La proporción de uso de tecnologías de información por la población infantil (de 10 a 15 años) es, en general, muy elevada. Por sexo, las diferencias en el uso de Internet son algo más significativas que en el uso de ordenador con 2,3 puntos y 1,5 respectivamente (a favor de las mujeres). La disponibilidad de teléfono móvil también se ha distanciado. En 2019 la disposición de móvil en las niñas superaba en 2,1 puntos a la de los niños. En 2020 ha aumentado a 3,5 puntos.

Al estudiar los tipos de actividades realizadas por Internet, se observa que tanto en hombres como en mujeres las actividades más realizadas son usar mensajería instantánea, telefonar o realizar video-llamadas a través de Internet, buscar información sobre bienes y servicios, recibir o enviar correos electrónicos y leer noticias, periódicos o revistas de actualidad *on-line* (INE, 2020).

Luego de analizar de manera general los datos que conforman la ETIC_H, a continuación, se describen cada una de las variables utilizadas en la presente investigación. Para ello se presentan los principales estadísticos descriptivos de las variables mostradas en el epígrafe 3.2. Primeramente, se exponen en la figura 2 los estadísticos descriptivos de la variable dependiente identificada a partir de las hipótesis planteadas en la investigación, así como las variables explicativas identificadas en la figura 3. Además, se relacionan la variable dependiente con las

independientes respecto a las hipótesis *H1*, *H2*, *H3* y *H4* establecidas en la investigación (figura 4, 5, 6 y 7 respectivamente).

Variable	Obs	Mean	Std. Dev.	Min	Max
USO_INT	15,343	.8001043	.3999348	0	1

Figura 2. Estadísticos descriptivos de la variable dependiente “USO_INT”. Fuente: elaboración propia en STATA.

Variable	Obs	Mean	Std. Dev.	Min	Max
NIVELEST	15,343	3.185752	2.273942	0	9
GROUP_EDAD	15,343	4.548915	1.788939	1	7
ING_HOG	15,343	3.13687	1.755164	1	6
SEXO	15,343	.4526494	.4977691	0	1
INTEFOR1	13,398	.8702791	.3778349	0	2
INTEFOR2	13,398	.8433348	.3681911	0	2
SIT_LAB	15,343	3.8615	2.580887	0	9
OCUPACION1	6,062	.2677334	.4428144	0	1
OCUPACION2	6,062	.0361267	.1866207	0	1
ACTIV	6,062	6.079017	3.064805	1	11

Figura 3. Estadísticos descriptivos de las variables independientes identificadas a partir de las hipótesis planteadas en la investigación. Fuente: elaboración propia en STATA.

NIVELEST	USO_INT		Total
	0	1	
0	1,052	361	1,413
1	1,302	1,394	2,696
2	435	2,732	3,167
3	168	2,901	3,069
4	4	20	24
5	34	1,468	1,502
6	41	1,433	1,474
7	16	1,812	1,828
8	1	133	134
9	14	22	36
Total	3,067	12,276	15,343

Figura 4. Relación entre variables identificadas (nivel de estudios y servicios usados de Internet) a partir de la *H1*.

Fuente: elaboración propia en STATA.

GROUP_EDAD	USO_INT		Total
	0	1	
1	1	1,053	1,054
2	5	1,095	1,100
3	11	2,324	2,335
4	62	2,642	2,704
5	299	2,604	2,903
6	726	1,799	2,525
7	1,963	759	2,722
Total	3,067	12,276	15,343

Figura 5. Relación entre variables identificadas (grupo de edad y servicios usados de Internet) a partir de la *H2*.

Fuente: elaboración propia en STATA.

ING_HOG	USO_INT		Total
	0	1	
1	1,164	1,621	2,785
2	953	3,579	4,532
3	230	2,606	2,836
4	36	968	1,004
5	43	1,317	1,360
6	641	2,185	2,826
Total	3,067	12,276	15,343

Figura 6. Relación entre variables identificadas (ingreso neto del hogar y servicios usados de Internet) a partir de la

H3. Fuente: elaboración propia en STATA.

SEXO	USO_INT		Total
	0	1	
0	1,889	6,509	8,398
1	1,178	5,767	6,945
Total	3,067	12,276	15,343

Figura 7. Relación entre variables identificadas (sexo y servicios usados de Internet) a partir de la *H4*. Fuente:

elaboración propia en STATA.

Atendiendo a la hipótesis 1 (*H1*), se observa en la figura 4 que el nivel de estudios tres (segunda etapa de la Educación Secundaria y similar) es el que tiene mayor demanda (2901) del uso del Internet. Por otro lado, haciendo alusión a la hipótesis 2 (*H2*), en la figura 5 se observa que los grupos de edades 3, 4 y 5 (de 35 a 64 años de edad) son las personas que tienen mayor uso del Internet con respecto a las demás. Además, en cuanto a la hipótesis 3 (*H3*), se observa en la

figura 6 que las personas con un ingreso de tipo dos (entre 900 y 1600 euros) son las que más hacen uso del Internet (3579). Así como, refiriéndose a la hipótesis 4 (*H4*), la figura 7 muestra que las mujeres (sexo = 0) hacen mayor uso (6509) del Internet.

A partir del análisis de los estadísticos descriptivos, así como la relación entre algunas variables implicadas en las hipótesis planteadas en la investigación, se tiene un mayor entendimiento de las observaciones que se tomaron como muestra de análisis. No obstante, sería provechoso también conocer, visualizar y agrupar la relación entre dos o más variables independientes, así como la asociación que existe entre variables independientes y dependientes.

Para ello, se debe de tener como punto de partida cómo funciona la correlación, cómo se interpreta la misma y sobre todo cómo se interpretan sus coeficientes. La correlación cuantifica cómo de relacionadas están dos variables sin considerar dependencias. A nivel experimental, la correlación se suele emplear cuando ninguna de las variables se ha controlado, simplemente se han medido ambas y se desea saber si están relacionadas. Primero se analiza si ambas variables están correlacionadas y, en caso de no estarlo, se procede a generar el modelo econométrico a utilizar.

El análisis de correlación es el primer paso para construir modelos explicativos y predictivos más complejos. A menudo es de interés observar y medir la relación entre dos variables numéricas mediante el análisis de correlación. Se trata de una de las técnicas más habituales en análisis de datos. La correlación puede ser entre dos variables exógenas, la cual es llamada como correlación parcial o entre una variable dependiente y una o más variables independientes, la cual es denominada correlación múltiple.

Coefficiente de correlación y su interpretación

El Coeficiente de correlación es un valor cuantitativo de la relación entre dos o más variables permitiendo medir el signo y magnitud de la tendencia entre las mismas. La magnitud que indica la fuerza de la relación puede variar desde -1.00 hasta 1.00. La correlación de proporcionalidad directa o positiva se establece con los valores +1.00 y de proporcionalidad inversa o negativa, con -1.00. Cuanto más cercano sea el valor a los extremos del intervalo (1 o -1) más fuerte será la tendencia de las variables, o será menor la dispersión que existe en los

puntos alrededor de dicha tendencia. Cuanto más cerca del cero esté el coeficiente de correlación, más débil será la tendencia, es decir, habrá más dispersión en la nube de puntos.

Si la correlación vale 1 o -1 se puede asegurar que la correlación es “perfecta”, si la correlación vale 0 se puede afirmar que las variables no están correlacionadas. En este caso lo ideal cuando se realiza una correlación parcial es que no exista relación entre las variables, lo cual ocurre cuando el coeficiente es de 0.00 o cercano. Siendo lo anterior la base para cuando se lleva a cabo la correlación parcial. Pero en el caso de correlación múltiple el valor de un coeficiente de correlación múltiple (R), se encuentra entre cero y uno. Cuanto más se acerque a uno mayor es el grado de asociación entre las variables. Y cuanto más se acerca a 0 la relación es peor (Chitarroni, 2002).

Como ya se ha afirmado, el coeficiente de correlación múltiple como se muestra entre las figuras 8 y la 13 se aplica a la situación en que una variable, a la que se puede llamar Y, ha sido aislada para examinar su relación con el conjunto de las otras variables.

```
(obs=15,343)
```

	USO_INT	NIVELEST
USO_INT	1.0000	
NIVELEST	0.4599	1.0000

Figura 8. Correlación múltiple entre la variable dependiente (uso del Internet) e independiente (nivel de educación alcanzado) asociadas a la H1. Fuente: elaboración propia en STATA.

```
(obs=15,343)
```

	USO_INT	GROUP_EDAD
USO_INT	1.0000	
GROUP_EDAD	-0.5405	1.0000

Figura 9. Correlación múltiple entre la variable dependiente (uso del Internet) e independiente (grupos de edad) asociadas a la H2. Fuente: elaboración propia en STATA.

```
(obs=15,343)
```

	USO_INT	ING_HOG
USO_INT	1.0000	
ING_HOG	0.1537	1.0000

Figura 10. Correlación múltiple entre la variable dependiente (uso del Internet) e independiente (ingresos netos del hogar) asociadas a la *H3*. Fuente: elaboración propia en STATA.

```
(obs=15,343)
```

	USO_INT	SEXO
USO_INT	1.0000	
SEXO	0.0688	1.0000

Figura 11. Correlación múltiple entre la variable dependiente (uso del Internet) e independiente (sexo) asociadas a la *H4*. Fuente: elaboración propia en STATA.

```
(obs=13,398)
```

	USO_INT	INTEFOR1	INTEFOR2
USO_INT	1.0000		
INTEFOR1	0.0823	1.0000	
INTEFOR2	0.0054	-0.1418	1.0000

Figura 12. Correlación múltiple entre la variable dependiente (uso del Internet) e independientes (formas de conexión a internet: banda ancha y móvil respectivamente) asociadas a la *H5*. Fuente: elaboración propia en STATA.

```
(obs=6,062)
```

	USO_INT	SIT_LAB	OCUPAC~1	OCUPAC~2	ACTIV
USO_INT	1.0000				
SIT_LAB	-0.0570	1.0000			
OCUPACION1	-0.1276	0.0849	1.0000		
OCUPACION2	0.0239	-0.0462	-0.1151	1.0000	
ACTIV	0.0594	-0.1472	-0.3548	-0.0168	1.0000

Figura 13. Correlación múltiple entre la variable dependiente (uso del Internet) e independientes (situación laboral, ocupación principal y actividad del establecimiento) asociadas a la *H6*. Fuente: elaboración propia en STATA.

Luego de arrojar y analizar los resultados que brinda realizar las correlaciones anteriormente presentadas, se puede observar la relación que existe entre las diferentes variables en cuestión. Se puede concluir que en el caso de las correlaciones realizadas se evidencia que entre las variables independientes existe poca o no relación. Teniendo como precedente lo anteriormente presentado

lo cual es aceptable, se puede proseguir en la investigación con las variables seleccionadas e implicadas en las hipótesis. Lo anterior da paso a poder llevar a cabo la implementación de los diferentes modelos Logit permitiendo lograr así el objetivo principal del trabajo.

3.4. Modelo

Como se indicó en el epígrafe 2.1, en el presente trabajo se pretenden realizar estimaciones mediante el modelo de elección discreta binario, específicamente el modelo Logit, con el objetivo de contrastar las hipótesis planteadas en la introducción y a su vez alcanzar el objetivo principal de la investigación. Teniendo en cuenta esto, seguidamente se explican las bases teóricas de este modelo y se plantean a partir de las hipótesis.

3.4.1. Modelo de elección discreta binaria

Según García existen varios economistas que están interesados en la utilización de los modelos de elección discreta binaria para explicar comportamientos observados. Para ello, necesitan estimar los parámetros de estos modelos utilizando una muestra de datos. Sin embargo, excepto el modelo de probabilidad lineal, los modelos de elección binaria son fundamentalmente no-lineales en parámetros, lo cual inmediatamente excluye el método de mínimos cuadrados ordinarios como estrategia de estimación. Este problema se resuelve utilizando máxima verosimilitud como método de estimación. Esta técnica se centra en la probabilidad de observar ciertas realizaciones de las decisiones objeto de estudio y las características de los individuos que las toman (García & Econometría, 2008).

En econometría es imprescindible conocer la naturaleza de las variables dependientes que se van a observar a partir de estimaciones. El modelo de elección discreta binaria es el modelo que más se ajusta en el caso de que la variable dependiente sea dicotómica la cual representa la decisión de realizar o no una acción determinada. Este modelo puede ser escrito como se muestra en la ecuación 1 (Domínguez & Navarro, 2019).

$$\mathbf{y}_i^* = \mathbf{x}_i' \boldsymbol{\beta} + \mathbf{u}_i$$

(1)

Donde \mathbf{y}_i^* es una variable latente (no observada) que se puede interpretar como un beneficio diferencial entre tomar una opción u otra. Para el individuo $i=1 \dots N$, $\mathbf{y}_i=1$ define la decisión de “éxito”, así como $\mathbf{y}_i=0$ sería en caso contrario “fracaso”. Por su parte, \mathbf{x}_i' es el vector de variables

independientes del modelo, β es el vector de coeficientes y u_i es el término de error aleatorio. Siguiendo las particularidades de la presente investigación y como se mencionó en los epígrafes anteriores en este trabajo se empleará el modelo Logit dentro del modelo de elección discreta binaria, el cual sigue una función de distribución logística. El modelo Logit puede expresarse como la ecuación 2.

$$\text{Prob}(y = 1) = \frac{e^{x_i'\beta}}{1 + e^{x_i'\beta}} \quad (2)$$

Es válido conocer que: si el parámetro β_j asociado a la variable x_j es positivo/negativo, entonces la probabilidad condicional $\text{Pr}(y_i = 1 | x_i; \beta)$ aumentará/disminuirá cuando x_j crece. No obstante, generalmente en estos estudios no solo interesa el signo del efecto sino también su magnitud. Con respecto a lo anterior, la medida que más se utiliza en estos modelos es el efecto marginal sobre la probabilidad condicional (García & Econometría, 2008). En este sentido, el efecto marginal para el modelo Logit se calcula mediante la ecuación 3.

$$f(x_i'\beta)\beta_j$$

Una vez detallado las bases teóricas del modelo de elección discreta binaria y específicamente el modelo Logit que se emplea en la presente investigación. Seguidamente se exponen estas bases teóricas mediante los modelos de estimación creados a partir de las hipótesis planteadas en el epígrafe 1, permitiendo alcanzar el objetivo principal de la investigación.

Hecha la observación anterior, se propone el *Modelo 1* correspondiente a la *H1* de la cual se describen en la tabla 6 las variables que intervienen; así como el *Modelo 2* correspondiente a *H2* de la cual se describen en la tabla 7 las variables que intervienen; así mismo, el *Modelo 3* que corresponde a la *H3* de la cual se describen en la tabla 8 las variables que intervienen. También se propone el *Modelo 4* correspondiente a la *H4* de la cual se describen en la tabla 9 las variables que intervienen; así como el *Modelo 5* y *Modelo 6* correspondientes a la *H5* y *H6* son descritas sus variables en las tablas 10 y 11 respectivamente.

Tabla 6. Descripción de las variables del *Modelo 1*.

Variable	Descripción	Tipo y relación
USO_INT (y)	Variable que incide en la decisión del uso del Internet.	Variable dependiente. Es la variable de interés para conocer si se usan o no los servicios de Internet.
NIVELEST (x)	Nivel de los estudios terminados	Variable independiente. Se relaciona con la variable dependiente por influir en el uso o no de los servicios de Internet.

Fuente: elaboración propia a partir de ETIC_H del INE y la H1.

Tabla 7. Descripción de las variables del *Modelo 2*.

Variable	Descripción	Tipo y relación
USO_INT (y)	Variable que incide en la decisión del uso del Internet.	Variable dependiente. Es la variable de interés para conocer si se usan o no los servicios de Internet.
GROUP_EDAD (x)	Variable que agrupa por grupos de edades	Variable independiente. Se relaciona con la variable dependiente por influir en el uso o no de los servicios de Internet.

Fuente: elaboración propia a partir de ETIC_H del INE y la H2.

Tabla 8. Descripción de las variables del *Modelo 3*.

Variable	Descripción	Tipo y relación
USO_INT (y)	Variable que incide en la decisión del uso del Internet.	Variable dependiente. Es la variable de interés para conocer si se usan o no los servicios de Internet.
ING_HOG (x)	Variable que representa el ingreso neto del hogar.	Variable independiente. Se relaciona con la variable dependiente por influir en el uso o no de los servicios de Internet.

Fuente: elaboración propia a partir de ETIC_H del INE y la H3.

Tabla 9. Descripción de las variables del *Modelo 4*.

Variable	Descripción	Tipo y relación
USO_INT (y)	Variable que incide en la decisión del uso del Internet.	Variable dependiente. Es la variable de interés para conocer si se usan o no los servicios de Internet.
SEXO (x)	Variable que representa el sexo de la persona encuestada	Variable independiente. Se relaciona con la variable dependiente por influir en el uso o no de los servicios de Internet.

Fuente: elaboración propia a partir de ETIC_H del INE y la H4.

Tabla 10. Descripción de las variables del *Modelo 5*.

Variable	Descripción	Tipo y relación
USO_INT (y)	Variable que incide en la decisión del uso del Internet.	Variable dependiente. Es la variable de interés para conocer si se usan o no los servicios de Internet.
INTEFOR1 (x_1)	Variable que incide en el tipo de conexión fija a Internet mediante banda ancha.	Variables independientes. Se relaciona con la variable dependiente por influir en el uso o no de los servicios de Internet.
INTEFOR1 (x_2)	Variable que incide en el tipo de conexión móvil a Internet mediante banda ancha.	

Fuente: elaboración propia a partir de ETIC_H del INE y la H5.

Tabla 11. Descripción de las variables del *Modelo 6*.

Variable	Descripción	Tipo y relación
USO_INT (y)	Variable que incide en la decisión del uso del Internet.	Variable dependiente. Es la variable de interés para conocer si se usan o no los servicios de Internet.
SIT_LAB (x_1)	Variable que incide en la situación laboral en la que se encuentra la persona.	Variables independientes. Se relaciona con la variable dependiente por influir en el uso o no de los servicios de Internet.
OCUPACION1 (x_2)	Variable que incide en si la ocupación principal es manual.	
OCUPACION2 (x_3)	Variable que incide en si la ocupación principal es del sector de las TIC.	
ACTIV (x_4)	Variable que incide en la actividad del establecimiento.	

Fuente: elaboración propia a partir de ETIC_H del INE y la H6.

4. Resultados

En este epígrafe se plasman las estimaciones de los modelos planteados a partir de los resultados obtenidos haciendo uso de STATA en su versión 16. Primeramente, se exponen los resultados tras aplicar el modelo Logit teniendo en cuenta cada una de las hipótesis planteadas en la investigación y luego se presentan los resultados de los efectos marginales tras realizar las estimaciones de cada modelo.

4.1. Modelo Logit

Una vez planteados los modelos que se estimarán en el presente trabajo, en este epígrafe se exponen las estimaciones realizadas mediante el modelo Logit. A continuación, se recogen los resultados de dichas estimaciones realizadas en el programa STATA entre las figuras 14 y 19.

Se observa en la figura 14 que el *Modelo 1* estima una máxima verosimilitud de -5070.9423 para un total de 15343 observaciones y una bondad de ajuste de 0.3393. Se puede interpretar que la variable independiente nivel de estudios terminados (NIVELEST) es estadísticamente significativa y positiva.

Logistic regression		Number of obs	=	15,343	
Log likelihood = -5070.9423		LR chi2(9)	=	5209.03	
		Prob > chi2	=	0.0000	
		Pseudo R2	=	0.3393	
USO_INT	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
NIVELEST					
1	1.137846	.0721531	15.77	0.000	.9964288 1.279264
2	2.907014	.0799096	36.38	0.000	2.750394 3.063633
3	3.918417	.1000887	39.15	0.000	3.722247 4.114587
4	2.679008	.5511086	4.86	0.000	1.598855 3.759161
5	4.834866	.1838848	26.29	0.000	4.474459 5.195274
6	4.623524	.1697314	27.24	0.000	4.290856 4.956191
7	5.799168	.2584038	22.44	0.000	5.292706 6.30563
8	5.959919	1.005604	5.93	0.000	3.988972 7.930865
9	1.521556	.3472805	4.38	0.000	.8408982 2.202213
_cons	-1.06957	.0609972	-17.53	0.000	-1.189123 -.9500182

Figura 14. Estimación del modelo logit para el *Modelo 1*, donde interviene la variable independiente (nivel de estudios terminados) con la dependiente (uso del Internet). Fuente: elaboración propia en STATA.

Del mismo modo se observa en la figura 15 que el *Modelo 2* estima una máxima verosimilitud de -4493.6955 para un total de 15343 observaciones y una bondad de ajuste de 0.4145. Se puede interpretar que la variable independiente grupo de edad (GROUP_EDAD) es estadísticamente significativa parcialmente y negativa.

Logistic regression		Number of obs = 15,343				
Log likelihood = -4493.6955		LR chi2(6) = 6363.52				
		Prob > chi2 = 0.0000				
		Pseudo R2 = 0.4145				
USO_INT	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
GROUP_EDAD						
2	-1.570312	1.096288	-1.43	0.152	-3.718998	.5783736
3	-1.606234	1.045119	-1.54	0.124	-3.65463	.4421622
4	-3.207227	1.008683	-3.18	0.001	-5.18421	-1.230243
5	-4.795023	1.002329	-4.78	0.000	-6.759552	-2.830495
6	-6.051947	1.001433	-6.04	0.000	-8.01472	-4.089175
7	-7.909611	1.00138	-7.90	0.000	-9.87228	-5.946942
_cons	6.959384	1.000467	6.96	0.000	4.998504	8.920263

Figura 15. Estimación del modelo logit para el *Modelo 2*, donde interviene la variable independiente (grupos de edades) con la dependiente (uso del Internet). Fuente: elaboración propia en STATA.

Asimismo, se muestra en la figura 16 que el *Modelo 3* estima una máxima verosimilitud de -6880.9458 para un total de 15343 observaciones y una bondad de ajuste de 0.1035. Se puede interpretar que la variable independiente ingreso neto del hogar (ING_HOG) es estadísticamente significativa y positiva.

Logistic regression		Number of obs = 15,343				
Log likelihood = -6880.9458		LR chi2(5) = 1589.02				
		Prob > chi2 = 0.0000				
		Pseudo R2 = 0.1035				
USO_INT	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
ING_HOG						
2	.9920429	.0529598	18.73	0.000	.8882437	1.095842
3	2.096312	.0787881	26.61	0.000	1.94189	2.250733
4	2.960532	.1740312	17.01	0.000	2.619437	3.301627
5	3.090731	.1596594	19.36	0.000	2.777804	3.403657
6	.8951608	.0591079	15.14	0.000	.7793115	1.01101
_cons	.3311809	.0384189	8.62	0.000	.2558813	.4064805

Figura 16. Estimación del modelo logit para el *Modelo 3*, donde interviene la variable independiente (ingresos netos del hogar) con la dependiente (uso del Internet). Fuente: elaboración propia en STATA.

También, se observa en la figura 17 que el *Modelo 4* estima una máxima verosimilitud de -7638.7539 para un total de 15343 observaciones y una bondad de ajuste de 0.0048. Se puede interpretar que la variable independiente sexo (SEXO) es estadísticamente significativa y positiva.

Logistic regression		Number of obs	=	15,343	
Log likelihood = -7638.7539		LR chi2(1)	=	73.40	
		Prob > chi2	=	0.0000	
		Pseudo R2	=	0.0048	
USO_INT	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
1.SEXO	.3511957	.0412954	8.50	0.000	.2702582 .4321332
_cons	1.237138	.0261345	47.34	0.000	1.185915 1.288361

Figura 17. Estimación del modelo logit para el *Modelo 4*, donde interviene la variable independiente (sexo) con la dependiente (uso del Internet). Fuente: elaboración propia en STATA.

Además, se observa en la figura 18 que el *Modelo 5* estima una máxima verosimilitud de -3830.5251 para un total de 13375 observaciones y una bondad de ajuste de 0.0363. Se puede interpretar que las variables independientes tipos de conexión fija (INTERFOR1) y móvil (INTERFOR2) a Internet mediante banda ancha son estadísticamente significativa y parcialmente positiva.

Logistic regression		Number of obs	=	13,375	
Log likelihood = -3830.5251		LR chi2(3)	=	288.68	
		Prob > chi2	=	0.0000	
		Pseudo R2	=	0.0363	
USO_INT	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
INTEFOR1					
1	1.15196	.0741006	15.55	0.000	1.006726 1.297195
2	-.6280205	.1728751	-3.63	0.000	-.9668494 -.2891915
INTEFOR2					
1	.5477462	.083623	6.55	0.000	.3838482 .7116442
2	0 (empty)				
_cons	1.015257	.1029722	9.86	0.000	.8134352 1.217079

Figura 18. Estimación del modelo logit para el *Modelo 5*, donde interviene la variable independiente (tipo de conexión fija y tipo de conexión móvil a Internet) con la dependiente (uso del Internet). Fuente: elaboración propia en STATA.

Se observa en la figura 19 que el *Modelo 6* estima una máxima verosimilitud de -423.44355 para un total de 5843 observaciones y una bondad de ajuste de 0.0972. Se puede interpretar que de las variables independientes situación laboral (SIT_LAB), ocupación principal manual (OCCUPATION1) y la actividad del establecimiento (ACTIV), esta última no es estadísticamente significativa, no obstante, las demás sí lo son, las cuales influyen negativamente sobre el uso del Internet. Cabe destacar que la variable referente a la ocupación principal sobre

las TIC (OCCUPATION2) se omite pues no cuenta con suficientes observaciones, lo cual no aporta mucha información excluyéndose en este estudio.

Logistic regression		Number of obs = 5,843				
Log likelihood = -423.44355		LR chi2(3) = 91.20	Prob > chi2 = 0.0000			
		Pseudo R2 = 0.0972				
USO_INT	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
SIT_LAB	-.4016333	.1239043	-3.24	0.001	-.6444813	-.1587854
OCUPACION1	-1.852022	.2546089	-7.27	0.000	-2.351047	-1.352998
OCUPACION2	0 (omitted)					
ACTIV	.0209875	.0363517	0.58	0.564	-.0502605	.0922355
_cons	5.60539	.3989083	14.05	0.000	4.823544	6.387236

Figura 19. Estimación del modelo logit para el *Modelo 6*, donde interviene la variable independiente (tipo de conexión fija y tipo de conexión móvil a Internet) con la dependiente (uso del Internet). Fuente: elaboración propia en STATA.

4.2. Efectos marginales

Una vez estimados los modelos Logit en este epígrafe se exponen los efectos marginales para cada modelo. Esto se puede observar entre las figuras 20 y 25. Mediante estas post-estimaciones se logra obtener interpretaciones más exactas sobre la probabilidad de hacer uso del Internet teniendo en cuenta las variables independientes en cada modelo.

Como se evidencia en la figura 20 se interpreta que, a menor nivel de estudio alcanzado por las personas, menor es la probabilidad de hacer uso del Internet. Tal es el caso de que al tener un nivel de estudios de tipo 0 (Analfabetos y estudios primarios incompletos) la probabilidad de uso de Internet correspondiente es de 0.25, mientras que al tener un nivel de estudios de tipo 8 (Doctorado) la probabilidad del uso de Internet correspondiente es de 0.99.

```
Adjusted predictions      Number of obs   =   15,343
Model VCE      : OIM
Expression     : Pr(USO_INT), predict()
```

	Delta-method				[95% Conf. Interval]	
	Margin	Std. Err.	z	P> z		
NIVELEST						
0	.2554848	.0116024	22.02	0.000	.2327445	.2782251
1	.5170623	.009624	53.73	0.000	.4981996	.5359251
2	.862646	.0061166	141.03	0.000	.8506576	.8746344
3	.945259	.0041061	230.21	0.000	.9372112	.9533069
4	.8333333	.0760726	10.95	0.000	.6842338	.9824328
5	.9773635	.0038379	254.66	0.000	.9698413	.9848857
6	.9721845	.0042832	226.98	0.000	.9637896	.9805795
7	.9912473	.0021786	455.00	0.000	.9869773	.9955172
8	.9925373	.0074348	133.50	0.000	.9779654	1.007109
9	.6111111	.0812497	7.52	0.000	.4518647	.7703575

Figura 20. Efecto marginal luego de estimar el *Modelo 1*. Fuente: elaboración propia en STATA.

La figura 21 describe los efectos marginales tras la post-estimación del *Modelo 2*, donde los grupos de edades 1, 2 y 3 (entre 16 y 44 años, haciendo referencia a las personas más jóvenes) son los que mayores probabilidades (0.999, 0.9954 y 0.9952 respectivamente) tienen de hacer uso del Internet con respecto a las demás edades. No obstante, partiendo de que al realizar la estimación para el *Modelo 2* como se evidenció en la figura 15, los grupos de edades 2 y 3 no eran significativos por lo que se rechaza la hipótesis planteada de que las personas jóvenes realizan mayor uso del Internet.

```
Adjusted predictions      Number of obs   =   15,343
Model VCE      : OIM
Expression     : Pr(USO_INT), predict()
```

	Delta-method				[95% Conf. Interval]	
	Margin	Std. Err.	z	P> z		
GROUP_EDAD						
1	.9990512	.0009483	1053.49	0.000	.9971925	1.00091
2	.9954545	.0020282	490.82	0.000	.9914794	.9994297
3	.9952891	.001417	702.37	0.000	.9925117	.9980664
4	.977071	.0028784	339.45	0.000	.9714294	.9827126
5	.8970031	.0056414	159.00	0.000	.8859462	.90806
6	.7124752	.0090072	79.10	0.000	.6948214	.7301291
7	.2788391	.0085951	32.44	0.000	.2619931	.2956851

Figura 21. Efecto marginal luego de estimar el *Modelo 2*. Fuente: elaboración propia en STATA.

Tras la estimación del *Modelo 3* mostrado en la figura 16, se evidenció que el ingreso del hogar (ING_HOG) con respecto al uso del Internet es significativo positivo. En ese mismo sentido, cabe agregar que tras la post-estimación se observan los efectos marginales en la figura 22, donde se observa que mientras aumenta el ingreso, mayor será la probabilidad de que se utilice el

Internet. Lo anterior puede apreciarse en los ingresos de tipo 3, 4 y 5 (entre 1600 euros o más) los cuales tienen probabilidades de 0.91, 0.964 y 0.968 de hacer uso del Internet.

```
Adjusted predictions      Number of obs   =   15,343
Model VCE      : OIM

Expression      : Pr(USO_INT), predict()
```

	Delta-method					[95% Conf. Interval]	
	Margin	Std. Err.	z	P> z			
ING_HOG							
1	.5820467	.0093461	62.28	0.000	.5637287	.6003647	
2	.7897176	.0060533	130.46	0.000	.7778533	.8015818	
3	.9188999	.0051262	179.26	0.000	.9088528	.9289469	
4	.9641434	.005868	164.31	0.000	.9526424	.9756444	
5	.9683824	.0047448	204.09	0.000	.9590827	.977682	
6	.7731776	.0078776	98.15	0.000	.7577377	.7886175	

Figura 22. Efecto marginal luego de estimar el *Modelo 3*. Fuente: elaboración propia en STATA.

Haciendo referencia a que se corroboró en la figura 17 la significancia positiva en la estimación del *Modelo 4*, en la figura 23 se observan los efectos marginales como post-estimación del modelo. En este sentido, se puede apreciar que los hombres (SEXO =1) tienen 0.83 de probabilidad de usar el Internet, en cambio las mujeres (SEXO = 0) tienen 0.77 de probabilidad. Por lo tanto, se observa claramente que los hombres hacen mayor uso del Internet que las mujeres.

```
Adjusted predictions      Number of obs   =   15,343
Model VCE      : OIM

Expression      : Pr(USO_INT), predict()
```

	Delta-method					[95% Conf. Interval]	
	Margin	Std. Err.	z	P> z			
SEXO							
0	.7750655	.0045563	170.11	0.000	.7661354	.7839956	
1	.8303816	.0045034	184.39	0.000	.8215551	.839208	

Figura 23. Efecto marginal luego de estimar el *Modelo 4*. Fuente: elaboración propia en STATA.

Siguiendo los resultados de significancia positiva tras la estimación del *Modelo 5* observados en la figura 18, donde además se aprecia que el hecho de que las personas no dieran una respuesta sobre la utilización del tipo de conexión fija de banda ancha a Internet incide negativamente en la decisión de uso del Internet. No obstante, tras realizar la post-estimación del modelo mediante los efectos marginales, en la figura 24 se observa que el hecho de utilizar el tipo de conexión fija (INTERFOR1 = 1) tiene una probabilidad de 0.931 de utilizar Internet en los hogares, sin

embargo, para el caso del tipo de conexión móvil (INTERFOR2 = 1) tiene una probabilidad de 0.918. Lo anterior demuestra que la probabilidad de que se use el Internet mediante conexión fija es mayor que mediante conexión móvil.

```
Predictive margins                                Number of obs   =   13,375
Model VCE      : OIM
Expression     : Pr(USO_INT), predict()
```

	Delta-method				
	Margin	Std. Err.	z	P> z	[95% Conf. Interval]
INTEFOR1					
0	.8120709	.0094624	85.82	0.000	.7935249 .8306168
1	.9314566	.0023682	393.32	0.000	.926815 .9360982
2	.698655	.0339534	20.58	0.000	.6321076 .7652025
INTEFOR2					
0	.8694256	.0081232	107.03	0.000	.8535044 .8853467
1	.9187463	.0025052	366.74	0.000	.9138362 .9236564

Figura 24. Efecto marginal luego de estimar el *Modelo 5*. Fuente: elaboración propia en STATA.

Haciendo alusión al *Modelo 6* estimado en la figura 19 se evidenció que todas las variables independientes resultaron ser significativas con influencia negativa con respecto a la variable dependiente uso del Internet (USO_INT), excepto las referentes a la actividad del establecimiento (ACTIV) y si la ocupación principal es TIC o no (OCCUPATION2) que no son significativas. No obstante, al realizar la post-estimación del modelo se ilustra en la figura 25 los efectos marginales del modelo. Mediante esto se puede apreciar que el hecho de que una persona tenga una situación laboral (SIT_LAB) del tipo 1, 2 o 3 (trabajando por cuenta ajena ya sea con contrato permanente o temporal, o por cuenta propia) con una probabilidad de 0.9852, 0.9829 y 0.9795 respectivamente, así como que su ocupación principal sea manual (OCCUPATION1) con una probabilidad de 0.9670, influyen negativamente sobre hacer uso del Internet.

```
Predictive margins                                Number of obs   =    5,495
Model VCE      : OIM

Expression     : Pr(USO_INT), predict()
```

	Delta-method				
	Margin	Std. Err.	z	P> z	[95% Conf. Interval]
SIT_LAB					
1	.9852479	.0021588	456.38	0.000	.9810167 .9894791
2	.9829633	.0043345	226.77	0.000	.9744678 .9914589
3	.9795302	.004037	242.64	0.000	.9716178 .9874426
OCUPACION1					
0	.9927742	.001491	665.85	0.000	.989852 .9956965
1	.9670029	.0050403	191.85	0.000	.9571241 .9768817
0.OCUPACION2					
	.9834395	.00169	581.91	0.000	.9801271 .9867519
ACTIV					
1	.9536079	.0101834	93.64	0.000	.9336489 .9735669
2	.9937869	.0023705	419.23	0.000	.9891408 .9984329
3	.9898184	.0041921	236.11	0.000	.981602 .9980348
4	.9824072	.004006	245.24	0.000	.9745556 .9902587
8	.986838	.0065277	151.18	0.000	.9740438 .9996321
9	.9898772	.0032683	302.87	0.000	.9834715 .9962829
10	.9780398	.0058084	168.38	0.000	.9666555 .989424
11	.9648413	.0339882	28.39	0.000	.8982255 1.031457

Figura 25. Efecto marginal luego de estimar el *Modelo 6*. Fuente: elaboración propia en STATA.

5. Conclusiones

Con el fin de dar cumplimiento al objetivo principal del presente trabajo se generaron las diferentes hipótesis que permitieron contrastar los factores determinantes del acceso y uso del Internet. Partiendo de ello se plantean diferentes modelos Logit por cada una de ellas, así como sus efectos marginales. Como resultado se aprecia que el nivel de educación alcanzado por los españoles influye positivamente sobre el hacer uso del Internet. En el caso del supuesto de que las personas jóvenes hacen mayor uso del Internet no se pudo interpretar pues el modelo arrojó que las edades comprendidas entre 16 y 44 años no resultaron ser estadísticamente significativas. En este sentido, se asume que el modelo habrá contemplado estos grupos de edades como no influyentes en la decisión del hacer uso del Internet, tal vez por las pocas personas que forman parte de estos grupos.

Por su parte, se comprueba que el ingreso neto del hogar influye positivamente en la decisión de hacer uso del Internet e incluso se puede afirmar que a medida que el ingreso en los hogares españoles aumenta su escala es directamente proporcional la probabilidad de hacer uso del Internet. Además, se constata que los hombres hacen mayor uso del Internet que las mujeres. Del mismo modo se comprueba que existe mayor probabilidad de hacer uso del Internet mediante la

conexión fija de banda ancha en los hogares españoles. También, se comprueba en el estudio que las personas con una situación laboral por cuenta ajena o cuenta propia, así como su ocupación principal sea manual influye negativamente sobre el hacer uso de Internet.

Por último, presentar que los elementos determinantes del acceso y uso de las TIC en los hogares españoles atendiendo a las variables analizadas en las diferentes hipótesis serían el nivel de estudio, el ingreso neto del hogar, la edad, el sexo, el tipo de conexión a Internet ya sea fija o móvil y dentro del estatus socioeconómico se identificaron los indicadores tales como situación laboral, así como la ocupación principal de tipo manual.

6. Referencias bibliográficas

- Alderete, M. V., & Formichella, M. M. (2020). Análisis Análisis de la primera brecha digital y su vínculo con el fracaso escolar en la Provincia de Buenos Aires: Asociación Argentina de Economía Política.
- Bordalba, M. M., & Bochaca, J. G. (2018). Accesibilidad y alfabetización digital: barreras para la integración de las TIC en la comunicación familia/escuela. *Revista de Investigación Educativa*, 36(1), 239-257.
- Catón Mayo, I., Cañón Rodríguez, R., & Grande de Prado, M. (2016). Brecha digital: impacto en el desarrollo social y personal. Factores asociados.
- Cuervo, M. R. V., & Menéndez, A. J. L. (2005). LA DIFUSIÓN DE LAS TIC EN ESPAÑA: UN ANÁLISIS DE LOS FACTORES DETERMINANTES.
- Chitarroni, H. (2002). El análisis de correlación y regresión lineal entre variables cuantitativas.
- Díaz-Vicario, A., Mercader Juan, C., & Gairín Sallán, J. (2019). Uso problemático de las TIC en adolescentes. *Revista electrónica de investigación educativa*, 21.
- Domínguez, M. M., & Navarro, D. G. (2019). Usos del teléfono inteligente en el sector rural de México. *Miradas críticas de la apropiación en América Latina*, 223.
- EL-DERECHO. (2020). Encuesta sobre Equipamiento y Uso de Tecnologías de Información y Comunicación en los Hogares. Año 2020. España (Instituto Nacional de Estadística – INE). from <https://elderecho.com/encuesta-sobre-equipamiento-y-uso-de-tecnologias-de-informacion-y-comunicacion-en-los-hogares-ano-2020-espana-instituto-nacional-de-estadistica-ine>
- Gairín Sallán, J., & Mercader, C. (2018). Usos y abusos de las TIC en los adolescentes.
- García, B., & Econometría, I. (2008). Modelos de elección binaria. *Econometría II*.
- INE. (2020). Instituto Nacional de Estadística. Encuesta sobre Equipamiento y Uso de Tecnologías de Información y Comunicación (TIC) en los Hogares, Base 2006-2020. from <https://www.ine.es/>
- Jiménez García, M., & Pérez Soto, F. (2020). Las TIC que influyen en un mayor uso del Internet en Oaxaca. *RIDE. Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 11(21).
- Mendoza-Ruano, J.-J., & Caldera-Serrano, J. (2014). Umbrales para la determinación de la brecha digital: comparativa entre regiones desarrolladas. *Transinformação*, 26, 125-132.
- Millán, J. M., Lyalkov, S., Burke, A., Millán, A., & van Stel, A. (2019). ‘Digital divide’ among European entrepreneurs: Which types benefit most from ICT implementation? *Journal of Business Research*.

- Moreira, M. A., Romero, O. C., & García, L. F. (2018). El uso escolar de las TIC desde la visión del alumnado de Educación Primaria, ESO y Bachillerato. *Educatio Siglo XXI*, 36(2 Jul-Oct), 229-254.
- Rodríguez Ruiz, J. G. (2019). Adopción de Internet en México: Propuesta de un índice con base en Microdatos. *Ensayos. Revista de economía*, 38(2), 135-182.
- Souza, C. A. d., Siqueira, É. S., & Reinhard, N. (2017). Digital divide of small and medium-sized enterprises: An analysis of influencing factors using the toe theory. *RAM. Revista de Administração Mackenzie*, 18, 15-48.
- Zamora Salas, G., & Rios Espinoza, G. L. (2019). ¿Conectar para incluir?: brecha digital en las personas con discapacidad. Análisis de su uso y apropiación de Internet desde un enfoque de capacidades.

7. Anexos

1. Estandarización de los datos a estimar.

```
* Normalizar la BD mediante comando infix, usando el fichero
proporcionado "ficherocuestionario".

infix CPRO 1-2 TOT_ENC 13-14 SEXO 19-19 EDAD 20-22 NACIONALIDAD
24-24 CONV 27-27 NIVELEST 29-29 SIT_LAB 30-30 ACTIV 32-33
OCUPACION1 34-34 OCUPACION2 35-35 ING_HOG 36-36 MUNICIPIO 37-37
TIP_H 41-41 TOT_MH 42-43 ORD 143-143 TABLET 144-144 TEL_FIJ 145-
145 TEL_MOV 146-146 VIV_INTER 147-147 INTEFOR1 148-148 INTEFOR2
149-149 USOMOV 190-190 USO_INT 191-191 ULT_INT 192-192 FREC_INT
193-193 USO_INT_VVD 194-194 SRV_INT_EMAIL 195-195
SRV_INT_VIDCALL 196-196 SRV_INT_REDSOC 197-197 SRV_INT_MSJ 198-
198 SRV_INT_SBIENSRV 199-199 SRV_INT_RDNEWS 200-200
SRV_INT_LISTMUSK 201-201 SRV_INT_TVCHANLS 202-202 SRV_INT_SEEMOV
203-203 SRV_INT_SEESITE 204-204 SRV_INT_GAME 205-205
SRV_INT_AONLINE 206-206 SRV_INT_INFOHEALTH 207-207
SRV_INT_APPMED 208-208 SRV_INT_FILEPHEALTH 209-209
SRV_INT_VBIENSRV 211-211 SRV_INT_ELECTBANK 212-212
SRV_INT_SHARECONT 213-213 SRV_INT_SAVFILE 214-214
SRV_INT_CSONLINE 215-215 SRV_INT_MAPREONLINE 216-216
SRV_INT_PROFALUMSITW 217-217 COMPRAS 228-228 VCOMPRAS 270-270
using ficherocuestionario
```

2. Modificación de variables originales de la base de datos.

```
* Para homogeneizar la BD se ajustan los datos de algunas
variables, si contiene el valor 6 se sustituye por 0, así como si
contiene valor 9 se sustituye por 2.
replace SEX0=0 if SEX0==6
replace OCUPACION1=0 if OCUPACION1==6
replace OCUPACION2=0 if OCUPACION2==6
replace INTEFOR1=0 if INTEFOR1==6
replace INTEFOR1=2 if INTEFOR1==9
replace INTEFOR2=0 if INTEFOR2==6
replace INTEFOR2=2 if INTEFOR2==9
replace USO_INT=0 if USO_INT==6

* Se crea la variable "Persona" para que exista en la BD una
columna que describa las personas, además se enumeran para tener
un control de las mismas.
sort CPRO
gen PERSONA =_n, before(CPRO)

* Se crean variables para separar grupos de edades en la BD.
gen GROUP_EDAD=1 if EDAD>=16 & EDAD<=24
replace GROUP_EDAD=2 if EDAD>=25 & EDAD<=34
replace GROUP_EDAD=3 if EDAD>=35 & EDAD<=44
replace GROUP_EDAD=4 if EDAD>=45 & EDAD<=54
replace GROUP_EDAD=5 if EDAD>=55 & EDAD<=64
replace GROUP_EDAD=6 if EDAD>=65 & EDAD<=74
replace GROUP_EDAD=7 if EDAD>=75
```