



**UNIVERSIDAD NACIONAL DEL ALTIPLANO**  
**FACULTAD DE INGENIERÍA ECONÓMICA**  
**ESCUELA PROFESIONAL DE INGENIERÍA ECONÓMICA**



**FACTORES QUE INCIDEN EN EL ACCESO Y USO DEL  
INTERNET Y SU IMPACTO SOBRE LA GENERACIÓN DE  
EMPLEO EN LOS HOGARES DEL PERÚ, AÑO 2020**

**TESIS**

**PRESENTADA POR:**

**BILL CLINTON CUTIPA MAMANI**

**PARA OPTAR EL TÍTULO PROFESIONAL DE:**

**INGENIERO ECONOMISTA**

**PUNO – PERÚ**

**2024**



## Reporte de similitud

NOMBRE DEL TRABAJO

**FACTORES QUE INCIDEN EN EL ACCESO  
Y USO DEL INTERNET Y SU IMPACTO SO  
BRE LA GENERACIÓN DE EMPLEO EN LO  
S HOGARES DEL PERÚ, AÑO 2020**

AUTOR

**BILL CLINTON CUTIPA MAMANI**

RECuento de palabras

**47976 Words**

RECuento de caracteres

**251046 Characters**

RECuento de páginas

**199 Pages**

Tamaño del archivo

**2.4MB**

Fecha de entrega

**Sep 30, 2024 9:24 AM GMT-5**

Fecha del informe

**Sep 30, 2024 9:26 AM GMT-5**

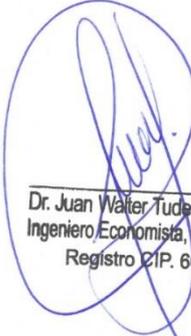
### ● 15% de similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para cada base de datos.

- 13% Base de datos de Internet
- Base de datos de Crossref
- 9% Base de datos de trabajos entregados
- 3% Base de datos de publicaciones
- Base de datos de contenido publicado de Crossref

### ● Excluir del Reporte de Similitud

- Material bibliográfico
- Material citado
- Material citado
- Coincidencia baja (menos de 12 palabras)

  
Dr. Juan Walter Tudela Mamani  
Ingeniero, Economista, Msc. Ph.D.  
Registro CIP. 66751



  
Dr. Sabino Edgar Mamani Choque  
Director de la Unidad de Investigación - FIE  
UNA - PUNO

Resumen



## DEDICATORIA

El presente trabajo de investigación quiero dedicar a aquellas personas que considero muy importantes en mi vida, debido a que les tengo mucho afecto y aprecio emocional.

En primer lugar, a dios, por derramar sus bendiciones sobre mí y llenarme de su fuerza para vencer todos los obstáculos que se me presentaron a lo largo de mi vida.

En segundo lugar, a mi familia, a mi padre Félix que está en el cielo, a mi madre Angélica y a mi hermano Rony, por haberme apoyado en todo momento de manera incondicional, por haberme forjado como la persona que soy, por haberme inculcado buenos valores y por haberme dado sus consejos, porque ellos han dado sentido a mi vida y todo lo que he logrado hoy, fue gracias a ellos.

En tercer lugar, a mis amigos, a mi amigo Eduard, a mi amigo Tomas y a mi amiga Edith, quienes sin esperar nada a cambio compartieron sus conocimientos, sus experiencias, sus tristezas y sus alegrías, asimismo, por motivarme, darme sus consejos y entregarme su sincera amistad.

Y por último, a todos mis compañeros y docentes de la Facultad de Ingeniera Económica, quienes durante los 5 años de mi formación profesional estuvieron a mi lado para enseñarme y apoyarme para que este sueño se haga realidad.

**Bill Clinton Cutipa Mamani**



## AGRADECIMIENTOS

Quiero expresar mi más sincero agradecimiento a todas las personas que han contribuido de alguna manera en la realización de esta tesis.

En primer lugar, quiero agradecer a la Universidad Nacional del altiplano y a mi querida Facultad de Ingeniería Económica, por haberme formado un profesional de calidad en el ámbito de lo académico y de la investigación.

En segundo lugar, agradecer a mi director/asesor de tesis, al Dr. Juan Walter Tudela Mamani, por su orientación, paciencia y apoyo constante a lo largo de todo el proceso de la presente investigación, asimismo, por sus valiosas sugerencias y comentarios que han sido fundamentales para lograr los objetivos propuestos.

Por último, también quiero agradecer a mis miembros de jurado, al Mg. Adalberto Calsin Sanchez, Dr. Fortunato Escobar Mamani y Dr. Manuel Timoteo Enriquez Tavera, por haberme apoyado a concluir de manera satisfactoria en todo el proceso de la ejecución y defensa de tesis.

**Bill Clinton Cutipa Mamani**



# ÍNDICE GENERAL

	Pág.
<b>DEDICATORIA</b>	
<b>AGRADECIMIENTOS</b>	
<b>ÍNDICE GENERAL</b>	
<b>ÍNDICE DE TABLAS</b>	
<b>ÍNDICE DE FIGURAS</b>	
<b>ÍNDICE DE ANEXOS</b>	
<b>ACRÓNIMOS</b>	
<b>RESUMEN .....</b>	<b>16</b>
<b>ABSTRACT.....</b>	<b>17</b>
<b>CAPÍTULO I</b>	
<b>INTRODUCCIÓN</b>	
<b>1.1. IDENTIFICACIÓN DEL PROBLEMA.....</b>	<b>18</b>
<b>1.2. FORMULACIÓN DEL PROBLEMA .....</b>	<b>20</b>
1.2.1. Problema general.....	20
1.2.2. Problemas específicos .....	20
<b>1.3. JUSTIFICACIÓN DE LA INVESTIGACIÓN .....</b>	<b>20</b>
1.3.1. Justificación teórica.....	20
1.3.2. Justificación práctica .....	21
<b>1.4. OBJETIVOS DE LA INVESTIGACIÓN.....</b>	<b>22</b>
1.4.1. Objetivo general .....	22
1.4.2. Objetivos específicos .....	22

## CAPÍTULO II

### REVISIÓN DE LITERATURA



<b>2.1.</b>	<b>ANTECEDENTES DE LA INVESTIGACIÓN .....</b>	<b>23</b>
2.1.1.	Antecedentes internacionales .....	23
2.1.2.	Antecedentes nacionales .....	35
2.1.3.	Antecedentes locales .....	40
<b>2.2.</b>	<b>MARCO TEÓRICO .....</b>	<b>42</b>
2.2.1.	Modelo de aceptación tecnológica (TAM) .....	42
2.2.2.	Teoría de los recursos y de la apropiación de nuevas tecnologías .....	44
2.2.2.1.	Motivación .....	49
2.2.2.2.	Acceso físico y material .....	49
2.2.2.3.	Habilidades digitales .....	50
2.2.2.4.	Uso de las TIC.....	51
2.2.3.	Teorías del empleo .....	52
2.2.3.1.	Teoría del empleo marxista.....	52
2.2.3.2.	La teoría del empleo keynesiana.....	52
2.2.3.3.	La teoría del empleo neoclásica .....	52
2.2.4.	Tipos de desempleo .....	52
2.2.5.	Teoría del bienestar .....	54
2.2.6.	Economía digital .....	55
2.2.7.	Teletrabajo.....	56
2.2.8.	Impacto de las TIC en el empleo.....	57
2.2.9.	Efectos de la pandemia de COVID-19 sobre la adopción de las TIC en el Perú .....	57
2.2.10.	Impacto del COVID-19 sobre el empleo en el Perú .....	58
2.2.10.1.	Impacto según tipo de trabajador .....	59
2.2.10.2.	Impacto según el tamaño de la empresa.....	60



2.2.10.3. Impacto según actividad económica .....	61
2.2.10.4. Impacto en el empleo juvenil .....	63
2.2.11. Modelos econométricos con variables dependientes discretas .....	63
2.2.11.1. Modelos de elección discreta binaria .....	63
2.2.11.2. Modelo de probabilidad no lineal (Logit y Probit) .....	66
2.2.11.2.1. Definición del modelo Logit .....	66
2.2.11.2.2. Definición del modelo Probit .....	69
2.2.12. Propensity score matching (PSM).....	75
2.2.12.1. Marco de evaluación y conceptos básicos de emparejamiento	75
2.2.12.2. Pasos para la implementación del emparejamiento por puntuación de propensión .....	80
2.2.12.3. Supuestos del emparejamiento por puntuación de propensión	87
<b>2.3. MARCO CONCEPTUAL .....</b>	<b>88</b>
2.3.1. Empleo .....	88
2.3.2. Empleo formal.....	88
2.3.3. Empleo informal.....	89
2.3.4. Desempleo.....	89
2.3.5. Población económicamente activa (PEA).....	89
2.3.6. Población económicamente inactiva (PEI) .....	90
2.3.7 PEA ocupada .....	90
2.3.8. PEA desocupada.....	90
2.3.9. Tasa de empleo.....	91
2.3.10. Tasa de desempleo abierto .....	91
2.3.11. Ingresos .....	91
2.3.12. Ingreso per cápita .....	91



2.3.13. Sociedad de la información y el conocimiento (SIC) .....	92
2.3.14. Brecha digital .....	92
2.3.15. Tecnologías de la información y la comunicación (TIC).....	92
2.3.16. Hardware .....	93
2.3.17. Software .....	93
2.3.18. Internet .....	93
2.3.19. Beneficios de internet.....	93
<b>2.4. HIPÓTESIS DE LA INVESTIGACIÓN .....</b>	<b>94</b>
2.4.1. Hipótesis general .....	94
2.4.2. Hipótesis específicas .....	94
<b>CAPÍTULO III</b>	
<b>MATERIALES Y MÉTODOS</b>	
<b>3.1. ENFOQUE DE INVESTIGACIÓN .....</b>	<b>96</b>
<b>3.2. ALCANCE DE INVESTIGACIÓN .....</b>	<b>96</b>
<b>3.3. DISEÑO DE INVESTIGACIÓN.....</b>	<b>96</b>
<b>3.4. TIPO DE INVESTIGACIÓN .....</b>	<b>97</b>
<b>3.5. MÉTODO DE INVESTIGACIÓN.....</b>	<b>97</b>
<b>3.6. UNIDAD DE ANÁLISIS .....</b>	<b>98</b>
<b>3.7. POBLACIÓN Y MUESTRA.....</b>	<b>98</b>
3.7.1. Población.....	98
3.7.2. Muestra.....	98
<b>3.8. TIPO DE MUESTREO .....</b>	<b>99</b>
<b>3.9. TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS.....</b>	<b>100</b>
3.9.1. Técnicas de recolección de datos .....	100
3.9.2. Instrumentos de recolección de datos .....	100



<b>3.10. FUENTES DE INFORMACIÓN Y PROCEDIMIENTO DE ANÁLISIS DE DATOS.....</b>	<b>100</b>
3.10.1. Fuentes de información .....	100
3.10.2. Procedimiento de análisis de datos .....	101
<b>3.11. IDENTIFICACIÓN Y DESCRIPCIÓN DE VARIABLES .....</b>	<b>102</b>
3.11.1. Descripción de las variables para el objetivo específico 1 .....	103
3.11.2. Descripción de las variables para el objetivo específico 2.....	104
3.11.3. Descripción de las variables para el objetivo específico 3.....	106
<b>3.12. ESPECIFICACIÓN DEL MODELO ECONOMETRICO .....</b>	<b>107</b>
3.12.1. Especificación del modelo econométrico para el objetivo específico 1	107
3.12.2. Especificación del modelo econométrico para el objetivo específico 2	109
3.12.3. Especificación del modelo econométrico para el objetivo específico 3	112
<b>3.13. OPERACIONALIZACIÓN DE VARIABLES .....</b>	<b>115</b>
3.13.1. Operacionalización de variables para el objetivo específico 1 .....	115
3.13.2. Operacionalización de variables para el objetivo específico 2 .....	116
3.13.3. Operacionalización de variables para el objetivo específico 3 .....	117

## **CAPÍTULO IV**

### **RESULTADOS Y DISCUSIÓN**

<b>4.1. RESULTADOS.....</b>	<b>120</b>
4.1.1. Presentación de las estadísticas descriptivas.....	120
4.1.1.1. Análisis descriptivo de las variables explicativas .....	120
4.1.1.2. Análisis descriptivo de la variable PEA.....	122
4.1.1.3. Análisis descriptivo de la variable acceso a internet.....	122
4.1.1.4. Análisis descriptivo de la variable uso de internet.....	124
4.1.1.5. Análisis descriptivo de la variable ingreso per cápita.....	125



4.1.2. Presentación de los resultados para el objetivo específico 1.....	127
4.1.2.1. Análisis del modelo inicial.....	127
4.1.2.2. Análisis del modelo final .....	129
4.1.2.3. Estimación de los efectos marginales .....	132
4.1.3. Presentación de los resultados para el objetivo específico 2.....	137
4.1.2.1. Análisis del modelo inicial.....	137
4.1.2.2. Análisis del modelo final .....	139
4.1.2.3. Estimación de los efectos marginales .....	142
4.1.4. Presentación de los resultados para el objetivo específico 3.....	148
4.1.4.1. Estimación de la puntuación de propensión.....	149
4.1.4.2. Elección de un algoritmo coincidente.....	151
4.1.4.3. Verificación del soporte común .....	151
4.1.4.4. Estimación del impacto.....	152
4.1.4.5. Prueba de la propiedad de balanceo .....	155
<b>4.2. DISCUSIÓN .....</b>	<b>157</b>
<b>V. CONCLUSIONES.....</b>	<b>164</b>
<b>VI. RECOMENDACIONES.....</b>	<b>167</b>
<b>VII. REFERENCIAS BIBLIOGRÁFICAS.....</b>	<b>169</b>
<b>ANEXOS.....</b>	<b>177</b>

**ÁREA:** Ciencias Económico Empresarial

**TEMA:** Empleo

**FECHA DE SUSTENTACIÓN:** 01 de octubre del año 2024



## ÍNDICE DE TABLAS

	<b>Pág.</b>
<b>Tabla 1</b> Descripción de las variables preliminares.....	102
<b>Tabla 2</b> Descripción de las variables definitivas para el O.E.1.....	104
<b>Tabla 3</b> Descripción de las variables definitivas para el O.E.2.....	105
<b>Tabla 4</b> Descripción de las variables definitivas para el O.E.3.....	106
<b>Tabla 5</b> Resumen de las variables para el O.E.1.....	116
<b>Tabla 6</b> Resumen de las variables para el O.E.2.....	117
<b>Tabla 7</b> Resumen de las variables para el O.E.3.....	118
<b>Tabla 8</b> Resumen de las estadísticas descriptivas .....	120
<b>Tabla 9</b> Conexión a internet en casa, por departamentos.....	123
<b>Tabla 10</b> Uso del servicio de internet, por departamentos .....	125
<b>Tabla 11</b> Ingreso neto per cápita, en términos de meses.....	126
<b>Tabla 12</b> Coeficientes de los modelos iniciales para el O.E.1.....	128
<b>Tabla 13</b> Coeficientes y medidas de ajuste de los modelos finales para el O.E.1. ...	129
<b>Tabla 14</b> Efectos marginales del modelo elegido para el O.E.1.....	132
<b>Tabla 15</b> Coeficientes de los modelos iniciales para el O.E.2.....	137
<b>Tabla 16</b> Coeficientes y medidas de ajuste de los modelos finales para el O.E.2. ...	139
<b>Tabla 17</b> Efectos marginales del modelo elegido para el O.E.2.....	142
<b>Tabla 18</b> Coeficientes y medidas de ajuste del modelo Logit del proceso PSM .....	149
<b>Tabla 19</b> Definición del soporte común.....	152
<b>Tabla 20</b> Efecto promedio de tratamiento en los tratados (ATT).....	152
<b>Tabla 21</b> Test de equilibrio .....	155



## ÍNDICE DE FIGURAS

	<b>Pág.</b>
<b>Figura 1</b> Modelo de aceptación tecnológica (TAM).....	43
<b>Figura 2</b> Modelo causal de recursos y teoría de la apropiación .....	45
<b>Figura 3</b> Etapas de acceso en la apropiación de la tecnología digital .....	48
<b>Figura 4</b> Reducción del empleo según tipo de trabajador, 2020 vs. 2019, a nivel de Lima Metropolitana .....	60
<b>Figura 5</b> Reducción del empleo según tamaño de empresa, 2020 vs. 2019, a nivel de Lima Metropolitana .....	61
<b>Figura 6</b> Reducción del empleo según actividad económica, 2020 vs. 2019, a nivel de Lima Metropolitana .....	63
<b>Figura 7</b> Población económicamente activa (PEA) .....	122
<b>Figura 8</b> Conexión a internet en casa .....	123
<b>Figura 9</b> Uso del servicio de internet .....	124



## ÍNDICE DE ANEXOS

	<b>Pág.</b>
<b>ANEXO 1.</b> Matriz de consistencia .....	177
<b>ANEXO 2.</b> Matriz de operacionalización de variables para el O.E.1. ....	179
<b>ANEXO 3.</b> Matriz de operacionalización de variables para el O.E.2. ....	180
<b>ANEXO 4.</b> Matriz de operacionalización de variables para el O.E.3. ....	181
<b>ANEXO 5.</b> Estimación del modelo inicial Logit O.E.1. ....	182
<b>ANEXO 6.</b> Estimación del modelo final Logit O.E.1. ....	183
<b>ANEXO 7.</b> Medidas de ajuste del modelo final Logit O.E.1. ....	183
<b>ANEXO 8.</b> Estimación del modelo inicial Probit O.E.1. ....	184
<b>ANEXO 9.</b> Estimación del modelo final Probit O.E.1. ....	185
<b>ANEXO 10.</b> Medidas de ajuste del modelo final Probit O.E.1. ....	185
<b>ANEXO 11.</b> Efectos marginales del modelo elegido O.E.1. ....	186
<b>ANEXO 12.</b> Efectos marginales del modelo elegido (sierra=0 y selva=0) O.E.1. ....	186
<b>ANEXO 13.</b> Estimación del modelo inicial Logit O.E.2. ....	187
<b>ANEXO 14.</b> Estimación del modelo final Logit O.E.2. ....	188
<b>ANEXO 15.</b> Medidas de ajuste del modelo final Logit O.E.2. ....	188
<b>ANEXO 16.</b> Estimación del modelo inicial Probit O.E.2. ....	189
<b>ANEXO 17.</b> Estimación del modelo final Probit O.E.2. ....	190
<b>ANEXO 18.</b> Medidas de ajuste del modelo final Probit O.E.2. ....	190
<b>ANEXO 19.</b> Efectos marginales del modelo elegido O.E.2. ....	191
<b>ANEXO 20.</b> Efectos marginales del modelo elegido (primaria=0, secundaria=0, sup_no_univ=0, sup_univ=0 y maestria=0) O.E.2. ....	192
<b>ANEXO 21.</b> Efectos marginales del modelo elegido (sierra=0 y selva=0) O.E.2. ....	193
<b>ANEXO 22.</b> Estimación de la puntuación de propensión O.E.3. ....	194



<b>ANEXO 23.</b> Estimación del impacto (vecino más cercano) O.E.3.....	195
<b>ANEXO 24.</b> Prueba de la propiedad de balanceo (vecino más cercano) O.E.3.....	195
<b>ANEXO 25.</b> Estimación del impacto (radio y calibre) O.E.3. ....	196
<b>ANEXO 26.</b> Prueba de la propiedad de balanceo (radio y calibre) O.E.3. ....	196
<b>ANEXO 27.</b> Estimación del impacto (kernel) O.E.3. ....	197
<b>ANEXO 28.</b> Prueba de la propiedad de balanceo (kernel) O.E.3. ....	197
<b>ANEXO 29.</b> Declaración jurada de autenticidad de tesis .....	198
<b>ANEXO 30.</b> Autorización para el depósito de tesis en el Repositorio Institucional....	199



## ACRÓNIMOS

ALADI:	Asociación Latinoamericana de Integración.
ATE:	Efecto Promedio del Tratamiento.
ATT:	Efecto Promedio del Tratamiento en los Tratados.
BCRP:	Banco Central de Reserva del Perú.
CEPAL:	Comisión Económica para América Latina y el Caribe.
CIA:	Supuesto de Independencia Condicional.
CIUS:	Encuesta sobre el Uso de Internet en Canadá.
ENAHO:	Encuesta Nacional de Hogares.
ENCV:	Encuesta de Calidad de Vida.
FITEL:	Fondo de Inversión en Telecomunicaciones.
IFLS:	Encuesta de Vida Familiar de Indonesia.
INEI:	Instituto Nacional de Estadística e Informática.
MIDIS:	Ministerio de Desarrollo e Inclusión Social.
MINEDU:	Ministerio de Educación.
MTC:	Ministerio de Transportes y Comunicaciones.
MTPE:	Ministerio de Trabajo y Promoción del Empleo.
OIT:	Organización Internacional del Trabajo.
OSIPTEL:	Organismo Supervisor de Inversión Privada en Telecomunicaciones.
PEA:	Población Económicamente Activa.
PEI:	Población Económicamente Inactiva.
PSM:	Emparejamiento por Puntuación de Propensión.
SIC:	Sociedad de la Información y el Conocimiento.
SIC:	Sociedad de la Información y Sociedad del Conocimiento.
TAM:	Modelo de Aceptación Tecnológica.
TIC:	Tecnologías de la Información y la Comunicación.
UIT:	Unión Internacional de Telecomunicaciones.



## RESUMEN

El objetivo general de la presente investigación fue analizar los factores que inciden en el acceso y uso de internet, y medir el impacto del uso de internet sobre la generación del empleo en los hogares del Perú, en el año 2020, utilizando la información de la Encuesta Nacional de Hogares (ENAHOG). Para tal fin, se consideró una muestra de 26,663 hogares del Perú y se planteó una investigación de enfoque cuantitativo, de alcance explicativo, de diseño cuasiexperimental y de tipo aplicada, debido a que para estimar la probabilidad de que un hogar tenga acceso a internet disponible y la probabilidad de que un jefe de hogar use el internet se utilizó los modelos de probabilidad no lineal (Logit y Probit), conjuntamente con el método de regresión econométrica de la máxima verosimilitud, a fin de estimar los coeficientes de cada covariable, mientras que para estimar el impacto del uso de internet sobre el empleo, se utilizó el método de evaluación de impacto del Propensity Score Matching (PSM), conjuntamente con las métricas de emparejamiento del vecino más cercano, radio y calibre y kernel. Dentro de los resultados que se obtuvo se pudo mostrar que las covariables más importantes en la determinación de la probabilidad de que un hogar tenga acceso a internet disponible fueron el celular y la computadora, con probabilidades del 26.62% y 44.44%, respectivamente. Por otro parte, las covariables más importantes en la determinación de la probabilidad de que un jefe de hogar use el internet fueron las covariables relacionadas a la educación superior, tales como: superior no universitaria, superior universitaria y maestría y doctorado, con probabilidades del 67.72%, 79.83% y 93.18%, respectivamente. Además, a través de las métricas de emparejamiento del vecino más cercano, radio y calibre y kernel se obtuvo que los jefes de hogar que usaron el internet, en promedio, percibieron un 35.54%, 38.01% y 36.22% más ingreso per capital mensual, respectivamente, en comparación a los jefes de hogar que no usaron el internet.

**Palabras Clave:** Acceso a internet, Empleo, Población económicamente activa, Uso de internet.



## ABSTRACT

The general objective of this research was to analyze the factors that influence access to and use of the Internet, and to measure the impact of Internet use on job creation in Peruvian households in 2020, using information from the National Household Survey (ENAHO). For this purpose, a sample of 26,663 households in Peru was considered and a quantitative, explanatory, quasi-experimental and applied research was proposed, since to estimate the probability that a household has access to the Internet available and the probability that a head of household uses the Internet, non-linear probability models (Logit and Probit) were used, together with the maximum likelihood econometric regression method, in order to estimate the coefficients of each covariate, while to estimate the impact of Internet use on employment, the Propensity Score Matching (PSM) impact assessment method was used, together with the nearest neighbor matching metrics, radius and caliber and kernel. The results obtained showed that the most important covariates in determining the probability that a household has access to the Internet available were the cell phone and the computer, with probabilities of 26.62% and 44.44%, respectively. On the other hand, the most important covariates in determining the probability that a household head uses the Internet were the covariates related to higher education, such as: non-university higher, university higher, and master's and doctorate, with probabilities of 67.72%, 79.83%, and 93.18%, respectively. In addition, through the nearest neighbor matching, radius and caliber, and kernel metrics, it was obtained that household heads who used the Internet, on average, received 35.54%, 38.01%, and 36.22% more monthly per capita income, respectively, compared to household heads who did not use the Internet.

**Keywords:** Internet access, Employment, Economically active population, Internet use.



# CAPÍTULO I

## INTRODUCCIÓN

### 1.1. IDENTIFICACIÓN DEL PROBLEMA

A pesar de que en los últimos 10 años la brecha digital ha disminuido de manera significativa por la globalización mundial y las políticas que han implantado los gobiernos, sin embargo, actualmente aún persiste esta brecha en todos los países del mundo, ya sea en mayor magnitud en los países subdesarrollados o menor magnitud en países desarrollados. Tal es la situación de la brecha digital, que, en el año 2020, las personas que usaban internet a nivel mundial alcanzaron a 4,537 millones aproximadamente, que representa el 58.75% de la población mundial (UIT, 2022). Mientras que en América Latina y Caribe la cantidad de usuarios de internet ascendieron a 458 millones aproximadamente, lo cual representa el 70.38% de esta población (CEPAL, 2021). Por otro lado, en la república del Perú, solo el 68.8% de las personas usaban internet, de las cuales la tasa de uso del internet en el área urbano fue del 69.35%, mientras que en el área rural fue del 39.70% (OSIPTTEL, 2022). Respecto al acceso de internet, en el año 2020, el 81.95% de los hogares peruanos tenían acceso a internet fijo y/o móvil, de las cuales la tasa de acceso en la zona urbana fue del 84.55% y en la zona rural fue del 55.10% (OSIPTTEL, 2022).

Por otra parte, cabe mencionar que el desempleo es un problema mundial que surge a causa de las fluctuaciones económicas, avances tecnológicos, globalización y escasa educación y formación de las personas. En ese contexto, en el año 2020, la tasa de desempleo a nivel mundial fue del 6.5%, que asciende a 220 millones de personas aproximadamente (OIT, 2021b). Mientras que en América Latina y Caribe la tasa de



desempleo fue del 10.3%, que asciende a 30 millones de personas aproximadamente (OIT, 2021b). Por otro lado, la tasa de desempleo que se registró en el Perú fue del 7.4% (MTPE, 2022).

Las principales causas de la brecha digital se deben a un fenómeno multidimensional, entre otros, explicado por la situación de la tecnología en general y en especial las vinculadas a las Tecnologías de la Información y la Comunicación (TIC), la infraestructura física, las condiciones demográficas y geográficas, el ingreso de la población y su distribución, y el nivel educativo y las políticas públicas de acceso. Todos estos factores actúan en mayor o menor medida, en la determinación de dicha brecha (ALADI, 2003).

Si esta situación problemática (brecha digital) continuara presentándose en la misma magnitud, entonces los factores socioeconómicos en el Perú se verían afectados en el corto, mediano y largo plazo, así como: los ingresos de las familias, la tasa de desempleo, la incomunicación de los habitantes en tiempo real, el desarrollo personal eficiente y eficaz, las oportunidades laborales, entre otros, los cuales tendrían una repercusión negativa para el desarrollo del país.

En este contexto, es necesario determinar los factores que influyen en el acceso y uso de internet y analizar el impacto que tiene el uso de internet sobre la generación del empleo. Si se logra identificar los factores específicos que expliquen el acceso y uso de internet, los gobiernos de diferentes niveles y los gestores públicos podrán desarrollar estrategias (como la creación y la ampliación de infraestructura y redes de internet) y plantear políticas (como las capacitaciones, entrega de subsidios monetarios, etc.) a fin de reducir la brecha digital. En consecuencia, la población peruana tendrá mayores posibilidades de tener acceso a internet y hacer su uso, asimismo, estará más capacitada



sobre el manejo e importancia del uso de internet, lo cual a la población le permitirá a tener mayores posibilidades de conseguir un empleo al momento de ingresar al mercado de trabajo.

## **1.2. FORMULACIÓN DEL PROBLEMA**

Tomando en consideración todos los argumentos mencionados en los párrafos anteriores, se plantearon las siguientes interrogantes a ser respondidas, para el desarrollo del presente trabajo de investigación.

### **1.2.1. Problema general**

- ¿Cuáles son los factores que inciden en el acceso y uso de internet, y que impacto genera el uso de internet sobre la generación del empleo en los hogares del Perú, en el año 2020?

### **1.2.2. Problemas específicos**

- ¿Cuáles son los factores que influyen en la probabilidad de que un hogar tenga acceso a internet disponible?
- ¿Cuáles son los factores que influyen en la probabilidad de que un jefe de hogar use internet?
- ¿Qué impacto genera el uso de internet sobre el empleo?

## **1.3. JUSTIFICACIÓN DE LA INVESTIGACIÓN**

### **1.3.1. Justificación teórica**

La presente investigación es importante, porque permitió a conocer los principales factores que determinan el acceso a internet disponible y el uso de internet, asimismo, ayudó a conocer el impacto que genera el uso de internet sobre



la generación del empleo. En ese contexto, la investigación planteada aportó un conocimiento adicional al ámbito de la investigación académica y científica, ya que se comprobó la teoría de los recursos y apropiación de nuevas tecnologías. Además, los resultados de esta investigación generarán reflexión y debate académico en las futuras investigaciones.

### **1.3.2. Justificación práctica**

Cabe mencionar también, que los resultados obtenidos en la presente investigación ayudarán a los gobiernos, funcionarios y gestores públicos a tomar decisiones, desarrollar estrategias y plantear políticas para reducir la brecha digital referente al uso de internet mediante las capacitaciones y asistencia técnica, fortaleciendo de esta manera las habilidades y el conocimiento de las personas del Perú sobre el manejo de internet, además, estos resultados ayudarán a los gobiernos, funcionarios y gestores públicos a contrarrestar la brecha digital de acceso, ampliando la infraestructura de redes de internet y realizando subsidios monetarios a las familias para que tengan mayores posibilidades de acceder al servicio de internet. De esta manera, a través del acceso y uso de internet, las personas del Perú tendrán mayores posibilidades de conseguir un empleo al momento de ingresar al mercado laboral y generar mayores ingresos, lo que finalmente en el largo plazo se traducirá en una reducción de la pobreza y en el desarrollo del país.



## **1.4. OBJETIVOS DE LA INVESTIGACIÓN**

### **1.4.1. Objetivo general**

- Analizar los factores que inciden en el acceso y uso de internet, y medir el impacto del uso de internet sobre la generación del empleo en los hogares del Perú, en el año 2020.

### **1.4.2. Objetivos específicos**

- Identificar los factores que influyen en la probabilidad de que un hogar tenga acceso a internet disponible.
- Determinar los factores que influyen en la probabilidad de que un jefe de hogar use internet.
- Estimar el impacto que genera el uso de internet sobre el empleo.



## CAPÍTULO II

### REVISIÓN DE LITERATURA

#### 2.1. ANTECEDENTES DE LA INVESTIGACIÓN

##### 2.1.1. Antecedentes internacionales

Noce & McKeown (2007), estimaron un modelo logístico de los factores que influyen en el uso de internet en Canadá para el año 2005, para lo cual utilizaron los datos de la Encuesta sobre el uso de internet en Canadá (CIUS) a fin de explorar en qué medida las variables como la edad, el ingreso, la ubicación, el sexo, la educación, el idioma de la entrevista y el número de niños en el hogar afectan al uso de internet por parte de los individuos en Canadá. Los resultados de la investigación se evidenciaron a través de los odds ratio (razón de probabilidades), demostrando que todas las variables explicativas son significativas, a excepción de la variable número de niños, en este contexto, un aumento de \$1,000 en los ingresos familiares, las probabilidades de uso de internet aumentan en un 1%; cada año adicional de edad resulta en una reducción de aproximadamente el 6% en las probabilidades de uso de internet; referente a la ubicación, las probabilidades de usar internet son aproximadamente 1.44 veces mayores para alguien que vive en un área urbana; ser hombre, reduce las probabilidades relativas de usar internet en un 20%; en relación a la educación, las probabilidades de usar internet son aproximadamente 3 veces mayores para alguien que tiene algún tipo de educación postsecundaria que para alguien que tiene, como máximo, una educación secundaria; y el uso del idioma de la entrevista ilustra que las probabilidades de utilizar internet son 1.3 veces mayores



para alguien que se entrevistó en inglés. En conclusión, los autores identificaron que los factores determinantes en el uso de internet en Canadá son: el nivel educativo, la ubicación, los ingresos familiares, la edad, el sexo y la educación, mientras que la presencia de niños en el hogar no tuvo la influencia esperada en el uso de internet.

Chaudhuri et al. (2005), analizaron el impacto de una variedad de influencias socioeconómicas en la decisión de los hogares de pagar por el acceso básico a internet, utilizando un marco de elección discreta y haciendo uso de una encuesta única sobre el uso de internet entre los hogares estadounidenses. Los resultados de la investigación indicaron que los ingresos y la educación son los predictores más fuertes de la compra de internet. Asimismo, un mayor grado de logro educativo y el hecho de ser estudiante son predictores positivos y significativos de la suscripción a internet. Además, varias características demográficas también entran en juego: los afroamericanos y los hispanos tienen menos probabilidades de estar en línea que otras categorías raciales, y las personas que están casadas tienen más probabilidades de tener acceso a internet. Curiosamente, tener un trabajo no está asociado de manera significativa con tener acceso a la red y ser hombre parece estar débilmente asociado con una menor probabilidad de elegir el acceso a internet, *ceteris paribus*. Finalmente, el precio mensual del acceso a internet tiene un impacto estadísticamente significativo, pero efectivamente pequeño en el acceso a internet.

Al-Hammadany & Heshmati (2011), identificaron los determinantes clave del uso de internet en Irak, para lo cual estimaron los efectos marginales para cada variable explicativa a través del modelo logístico binomial, obteniendo de esta



manera los siguientes resultados: para la variable género, los hombres tienen una probabilidad mayor de tener internet en 0.154 respecto de las mujeres; referente a la edad, la categoría de 25 a 40 años tiene 0.011 menos probabilidad de usar internet en comparación con la categoría de 12 a 25 años y la categoría de mayores de 40 años tiene 0.135 menos probabilidad de usar internet en comparación con la categoría de 12 a 25 años; en relación a la educación, las categorías de escuela superior o preparatoria (high school), bachiller universitario en ciencias (B.Sc.), y magister en ciencias (M.Sc.) y doctorado (Ph.D.) tienen 0.146, 0.237 y 0.388 más probabilidades de tener internet y usarlo, respectivamente, en comparación con la categoría de escuela primaria (primary school) y secundaria (intermediate school); respecto al tipo de trabajo, el empleado del sector público, el empleado del sector privado y el estudiante tienen 0.134, 0.131 y 0.068 mayores probabilidades de tener y usar internet, sucesivamente, en comparación con los desempleados; para el propósito del uso de internet, los que utilizan el internet con el propósito especial tienen una mayor probabilidad de tener internet en 0.028 respecto de las personas que utilizan el internet de manera no especificada (búsqueda general), pero los que no son usuarios de internet (sin respuesta) tienen una menor probabilidad de usar internet en 0.770 respecto de los que utilizan el internet de manera no especificada; con referente al costo, tiene un efecto positivo y estadísticamente significativo, pero su impacto no es grande en comparación con otros factores. Finalmente, concluyeron que el nivel de educación es el factor más importante que influye en el uso de internet, además, otras características, como la edad, el género, el tipo de empleo y la tasa de utilización, también resultaron ser factores significativos en la adopción de internet, pero el nivel de costos



(asequibilidad) no tiene un efecto fuerte en la adopción de internet porque está ampliamente disponible para el sector público en Irak.

Siaw et al. (2020), analizaron los efectos del uso de internet en los ingresos agrícolas y familiares utilizando datos de encuestas realizadas a 478 agricultores rurales de dos regiones de Ghana, para lo cual se emplearon un modelo de regresión de conmutación endógena (ESR) y el modelo Probit. Los resultados de las variables de control sobre la variable de tratamiento revelaron lo siguiente: los agricultores que trabajan fuera de la granja tienen un 4.73% más de probabilidades de usar el internet en comparación de los que trabajan en la granja; los que tienen una educación secundaria o superior tienen un 6.14% más de probabilidades de usar el internet en comparación de los que tienen educación primaria o ningún nivel educativo; los que tienen un acceso a activos no fijos (NFA) tienen una probabilidad mayor de hacer el uso de internet, de un 0.12%, respecto de los que no tienen un acceso a activos no fijos; los que tienen acceso a crédito tienen un 6.22% más de probabilidades de hacer el uso del internet en comparación de los que no tienen este acceso; a medida que la edad se incrementa en un año, los agricultores tienen un 11.7% menos probabilidades de utilizar internet; a medida que el tamaño de las granjas de los agricultores aumentan en un acre tienen una menor probabilidad de hacer el uso de internet, de un 1.24%; los que perciben que navegar con un teléfono y/o computadora es fácil tienen un 2.23% más de probabilidades de utilizar internet. Mientras que los resultados de la variable de tratamiento sobre las variables de resultado, evidenciaron que el uso de internet aumentó los ingresos agrícolas y familiares en un 20.10% y un 15.47%, respectivamente. Finalmente, los autores llegaron a la conclusión de que el uso de internet se vio influenciado por el empleo fuera de la granja, la educación, el



acceso al crédito, los activos no fijos (NFA), la edad y las variables de percepción, además, el uso de internet aumentó los ingresos agrícolas y familiares.

Rini & Rahadiantino (2020), analizaron el impacto de la utilización de internet en el bienestar de los hogares entre dos grupos, empresas familiares con uso de internet para fines comerciales y sin uso de internet para fines comerciales, mediante la aplicación del método de evaluación de impacto del Propensity Score Matching (PSM) y empleando la regresión del Probit binomial. Para tal fin, utilizaron los datos de la Encuesta de Vida Familiar de Indonesia (IFLS) para el período de 2014. Los resultados mostraron que la probabilidad de que las empresas domésticas utilicen internet se ve afectada por el área de residencia, el tamaño del hogar, la edad del jefe de familia, el estado civil, el uso del teléfono móvil, el tipo de negocio, la antigüedad del negocio y los activos del negocio. También los resultados evidenciaron que el uso de internet para los negocios tiene impactos positivos en el bienestar de los hogares, es decir, utilizando el emparejamiento por el vecino más cercano, la utilización de internet por parte de las empresas aumenta el gasto per cápita en un 19.2%, con el emparejamiento por estratificación, aumenta el gasto per cápita en un 25.7% y utilizando el emparejamiento de kernel, aumenta el gasto per cápita en un 27.1%.

Priyatna (2022), analizó el efecto del uso de internet en los ingresos de los hogares en las zonas rurales de Indonesia utilizando datos de Sakernas 2018, para lo cual empleó el método de emparejamiento por puntaje de propensión (PSM). Los resultados que obtuvo mostraron que las probabilidades de usar internet son causadas por el tamaño de la familia, la edad del jefe de familia, la experiencia laboral, el estado civil, la educación formal e informal, la experiencia de graves



dificultades en las funciones corporales, el género, el sector empresarial y la gestión financiera, además, el uso de internet aumenta los ingresos de los hogares en las zonas rurales de Indonesia en un 29%.

Worgwu et al. (2023), examinaron el impacto del emprendimiento social y la utilización de las tecnologías de la información y la comunicación (TIC) en el empleo en Nigeria, para lo cual utilizaron los datos de la encuesta obtenida de YouWiN en el año 2019, empleando estadísticas descriptivas, mínimos cuadrados ordinarios (OLS) y emparejamiento por puntaje de propensión (PSM). Los resultados mostraron que el emprendimiento social contribuye aproximadamente en un 21% al empleo. Además, la utilización de las TIC por parte de las empresas, el acceso a internet y la comunicación con los clientes por correo electrónico contribuyen aproximadamente en un 66% y un 64%, respectivamente, a la capacidad de las empresas para contribuir a la creación de empleo.

Osabohien et al. (2022), aplicaron el método Propensity Score Matching (PSM) y Mínimos Cuadrados Ordinarios (OLS) para estimar el impacto de la innovación empresarial en el empleo futuro en Nigeria, utilizando la encuesta de referencia del emprendimiento juvenil con innovación (YouWiN) en el año 2019. Las medidas de innovación que consideraron fueron: la introducción de nuevos productos y servicios, el acceso a internet, la utilización de correos electrónicos para marketing y la participación de la empresa en la publicidad en línea. Los resultados que obtuvieron en relación a los efectos promedio del tratamiento en los tratados (ATT) tanto para el emparejamiento con el vecino más cercano (NNM) como para el emparejamiento basado en kernel (KBM), indican que las empresas que introducen nuevos productos y servicios contribuirían con un



46.98% al empleo futuro, las empresas que tienen acceso a internet contribuirían con un 54.65% al empleo futuro, las empresas que utilizan el correo electrónico para marketing contribuirían con un 41.87% al empleo futuro y las empresas que participan en la publicidad en línea contribuirían con un 69.64% al empleo futuro.

Mwirigi (2013), investigó los factores que influyen en la adopción de la conectividad a internet entre los residentes del municipio de Meru, mediante un estudio descriptivo y utilizando el método de muestreo aleatorio simple para seleccionar a los encuestados de entre varias personas conectadas en todo el municipio. De acuerdo a los resultados que obtuvo llegó a la conclusión de que el alto coste de la suscripción a internet fue identificado como uno de los principales factores que obstaculizan el acceso a internet por la mayoría de los jóvenes e incluso los adultos en el municipio de Meru, además, la publicidad fue identificada como el segundo factor influyente, y la mayoría de los encuestados opinaron que los proveedores de internet tienen mucho que hacer para llegar a los clientes potenciales, por otra parte, los niveles de alfabetización se identificaron como el tercer factor influyente y por último la velocidad de conexión fue identificada como el factor menos importante que puede atribuirse a la presencia de un área de paquetes de alta velocidad alrededor del municipio de Meru.

Haight et al. (2014), estimaron el impacto de los factores demográficos en el acceso a internet, en el nivel de actividad en línea y en el uso de sitios de redes sociales en la sociedad canadiense, basándose en la encuesta sobre el uso de internet en Canadá (CIUS) en el año 2010. Respecto al acceso a internet obtuvieron que los ingresos, la educación, el entorno (rural y urbano), el estado migratorio y la edad afectan su patrón de adopción, evidenciando de esta manera



que las familias que ganan entre \$25,000 y \$40,999, entre \$41,000 y \$65,000, entre \$65,000 y \$100,000 y entre \$100,000 y más, tienen 1.40, 2.17, 2.80 y 5.40 veces más probabilidades de acceder a internet, respectivamente, en comparación con las personas con ingresos familiares inferiores a \$25,000; las personas con niveles de educación más altos tienen más probabilidades de acceder a internet; las personas que viven en áreas urbanas tienen 1.51 veces más probabilidades de acceder a internet en comparación con las de áreas rurales; los inmigrantes nacidos en Canadá o que inmigraron antes de su nacimiento tienen 1.74 veces más probabilidades de acceder a internet en comparación con los inmigrantes recientes; a medida que aumenta la edad, la probabilidad de que una persona tenga acceso a internet disminuye drásticamente en 0.94 veces.

Zhang & Liu (2023), utilizaron el método de comparación por puntaje de propensión (PSM) y el modelo de doble diferencia (DID) para estimar el impacto del uso de internet en la innovación empresarial. Sus resultados mostraron que el uso de internet puede promover significativamente el nivel de innovación de las empresas manufactureras exportadoras de China, con una heterogeneidad obvia entre regiones. En comparación con las regiones central y occidental, el papel de internet en la promoción de la innovación de las empresas manufactureras exportadoras en la región oriental es más significativo. En comparación con las regiones no costeras, las empresas manufactureras exportadoras en las regiones costeras se benefician más. El efecto de promoción de internet en la innovación de las empresas manufactureras exportadoras tiene un umbral de escala y varía según el tipo de propiedad, es decir, es más significativo en las empresas grandes y medianas, mientras que las empresas estatales tienen un nivel de innovación más alto que las empresas privadas.



Grazzi (2011), evaluó los principales determinantes socioeconómicos del uso de internet incondicional y conexión a internet en los hogares de siete países de América Latina, utilizando los datos de las Encuestas Nacionales de Hogares de Brasil, Chile, México, Costa Rica, El Salvador, Honduras y Paraguay. Los resultados que obtuvo sobre el uso incondicional de internet, indican que los individuos mayores, menos educados y más pobres tienen menos probabilidades de utilizar internet en todos los países, por ejemplo, en Chile a medida que se incrementa el ingreso, la educación y la edad de los individuos, su probabilidad de utilizar internet aumenta en 9.3%, 5.1% y 1.1%, respectivamente. Además, en todos los países, excepto Honduras, hay evidencia estadística de que las mujeres tienen menos probabilidades de usar internet, por ejemplo, en Chile las mujeres tienen un 2% menos probabilidades de usar internet, respecto a los hombres. Por otra parte, las personas que viven en áreas urbanas tienen muchas más probabilidades de usar internet que las que viven en áreas rurales, por ejemplo, en Chile se evidencia que las personas que viven en hogares rurales, en promedio y ceteris paribus, tienen un 12.8% menos probabilidades de usar internet. Por otro lado, la condición de estudiante también es un determinante positivo y significativo del uso de internet en todos los países considerados, por ejemplo, en Chile un estudiante tiene un 58.2% más probabilidad de estar conectado a internet, en promedio y ceteris paribus, que no estudiantes. En cuanto a la situación laboral, el análisis empírico muestra resultados mixtos: en Brasil, Chile, Costa Rica y México los coeficientes estimados son positivos y significativos, en Paraguay son negativos y significativos, y en El Salvador y Honduras no son estadísticamente significativos. Por último, los coeficientes asociados a vivir en un hogar con conexión a internet disponible son positivos y significativos al 1% en todos los



países considerados, en otras palabras, cuando una persona tiene una conexión disponible a su disposición, su probabilidad de usar internet es considerablemente mayor en 3.94% para el caso de Chile. Mientras que los resultados que obtuvo sobre la conexión a internet en casa, muestran que los hogares con altos ingresos, presencia de mayores niveles educativos de los miembros, mayor densidad de estudiantes, mayor tamaño de familia tienen mayores probabilidades de tener conexión a internet en casa, mientras que los hogares que se encuentran en el área rural tienen menores probabilidades de tener conexión a internet en casa.

Botello (2014), utilizó la Encuesta de Calidad de Vida (ENCV) del Departamento Administrativo Nacional de Estadísticas (DANE) y un modelo de elección discreta para identificar los determinantes del acceso a internet en los hogares colombianos durante el año 2013, en donde obtuvo los siguientes resultados: un incremento del 1% en el ingreso mensual de las personas, incrementa su probabilidad de usar internet en un 2.62%; las personas ubicadas en la región central del país tienen más probabilidad de tener acceso a internet, cercanas al 20% en promedio; los habitantes de los centros urbanos poseen una probabilidad mayor del 10% de tener acceso a internet, comparativamente con los que se encuentran en el área rural de los municipios; los hombres muestran mayores probabilidades del acceso a internet, una mujer cuenta con una diferencia negativa del 10% en contra; si se da un aumento del 1% en las horas de cuidado familiar (niños o labores hogar), la probabilidad del acceso a internet se reduce un 4.2%; las personas que tienen títulos de posgrado, educación universitaria, educación secundaria y educación primaria tienen un 84%, 83%, 38% y 13% de posibilidades más de usar internet, respectivamente, que una persona que no tenga



ningún grado educativo; y un aumento de un año, disminuye un 1.15% la probabilidad de usar internet.

Navarro (2010), estimó el impacto del uso de internet en los ingresos individuales en seis países de América Latina, utilizando los datos transversales provenientes de las encuestas nacionales de hogares para Brasil, Chile, Costa Rica, Honduras, México y Paraguay. Dado los diferentes patrones de uso de internet y sus implicaciones, el análisis lo realizó por separado para los trabajadores asalariados y los autónomos. Mientras que los usuarios asalariados acceden a internet principalmente en el trabajo, los usuarios autónomos lo hacen principalmente en otros lugares. Para estimar el impacto, implementó el método de emparejamiento de la puntuación de propensión (PSM), conjuntamente con la métrica del vecino más cercano, considerando como variables de resultado el trabajo autónomo y asalariado, como variable de tratamiento el uso del internet y como variables de control la edad, el logro educativo, la ocupación, el sector de actividad, el tipo de casa, la propiedad de la casa, el acceso a televisión satelital, el acceso a un teléfono fijo y el acceso a una computadora en el hogar. Los resultados empíricos que obtuvo reflejaron un retorno considerable al uso de internet tanto para los trabajadores asalariados como para los trabajadores por cuenta propia, del 18% y el 30%, respectivamente.

Barrios (2021), utilizó los modelos de regresión Logit binomial y Logit multinomial para analizar los factores determinantes de acceso y de uso de internet de los hogares en cabeceras y centros poblados en Colombia, a partir de la Encuesta de Calidad de Vida del año 2019. Los resultados que obtuvo indican que la probabilidad de tener acceso a internet es menor para los hombres que para las



mujeres, por otra parte, la probabilidad promedio para jefes de hogar con educación primaria o sin educación es de 0.17, mientras que para los que tienen con postgrado es de 0.93, por el contrario, la probabilidad de acceder a internet disminuye a medida que aumenta la edad del jefe del hogar; para aquellos menores de 25 años la probabilidad es en promedio 0.67 y para aquellos mayores de 65 años es de 0.17. En ese contexto, el autor concluyó que los principales determinantes del acceso a internet son: el ingreso de los hogares, la edad, la educación y las habilidades en el uso de las TIC.

Martínez (2020), analizó las barreras de acceso y uso de las Tecnologías de la Información y Comunicaciones en las zonas rurales y urbanas de México. El análisis empírico consistió en explicar el no acceso y no uso de internet, condicionado sobre un conjunto de características demográficas y socioeconómicas, mediante regresiones logísticas multinomiales. Por lo tanto, los resultados que obtuvo evidencian que el índice de riqueza se correlaciona de forma positiva con la falta de interés. Además, un incremento del 1% en el índice de riqueza se asocia con un aumento de 10.2% y 12.5%, en el sector rural y urbano, respectivamente, de mencionar la falta de interés como la principal razón de no uso de internet. De igual forma, la edad es una barrera para el uso de la red, donde la magnitud del efecto es mayor para las habilidades digitales que la falta de interés, pues a mayor edad aumenta la probabilidad de señalar las competencias digitales como una barrera para el no uso, en 29.2% para el sector rural y 6% en el urbano. Con respecto a la educación, para los individuos con estudios universitarios y superiores, en comparación a aquellos sin educación, se reduce la posibilidad de contar con las habilidades digitales como la principal razón para no utilizar el internet en 79.8% en zonas rurales y 81.9% en espacios urbanos. Por



último, la presencia de niños en el hogar disminuye la probabilidad de que los residentes rurales y urbanos masculinos indiquen la falta de interés como motivo de no uso en 19.4% y 21.3%, respectivamente. Por lo tanto, concluyó que el costo, la edad, los bajos niveles de educación y la falta de habilidades digitales se relacionan con la no conectividad y el no uso de internet en contextos rurales y urbanos, aunque con diferentes magnitudes.

Lera et al. (2009), analizaron el impacto de variables sociodemográficas y regionales en la probabilidad de uso de internet en España a nivel individual para el año 2007, en base a la Encuesta sobre Equipamiento y Uso de Tecnologías de Información y Comunicación en los Hogares (TICH) elaborada por el Instituto Nacional de Estadística con periodicidad semestral desde el año 2005. A través de la regresión del modelo Probit obtuvieron que las variables sociodemográficas tales como la educación, la edad y la ocupación explican el uso de internet. Además, determinaron que las variables regionales, como el porcentaje de empleo en el sector servicios, el PIB per cápita y el capital neto en TIC, explican también una mayor probabilidad de uso de internet por los individuos.

### **2.1.2. Antecedentes nacionales**

Fernández & Medina (2011), evaluaron el impacto del acceso a las Tecnologías de Información y Comunicación (TIC) sobre el nivel de ingresos de los hogares del Perú. Para estimar dicho impacto utilizaron dos metodologías de evaluación: la metodología del Propensity Score Matching y la estimación de datos de panel con efectos fijos. Los resultados que obtuvieron evidencian que en cuanto a la edad de los miembros del hogar, para el acceso a las TIC en conjunto, el contar con niños menores a los 10 años disminuye la probabilidad de acceso en



3%, mientras que la presencia de adultos casados y jóvenes entre los 18 y 25 años aumenta dicha probabilidad en 2% y 1%, respectivamente, asimismo, la educación de los miembros del hogar juega un rol fundamental en el acceso a TIC, pues se observa que a mayores niveles de educación, la probabilidad del acceso aumenta en 3%. Además, los resultados que obtuvieron con el Propensity Score Matching indican que el acceso conjunto a estos medios puede incrementar el ingreso promedio per cápita del hogar en 105 nuevos soles. Más aún, el acceso a telefonía fija, telefonía móvil e internet puede aumentar el ingreso promedio per cápita del hogar en 19, 132 o 365 nuevos soles. Por otra parte, los resultados que obtuvieron mediante los datos de panel con efectos fijos muestran que la telefonía móvil y el internet presentan un efecto positivo sobre el nivel de ingresos, de 28 y 104 nuevos soles, respectivamente. Ambos servicios, de manera conjunta, aumentan el ingreso en 216 nuevos soles.

Ministerio de Transportes y Comunicaciones (MTC, 2020), estimó los impactos que tienen el acceso y uso de los servicios de internet sobre el bienestar de los hogares en el Perú, para el periodo 2017-2019, en base a los datos de panel de la Encuesta Nacional de Hogares (ENAHOG). Para ello, empleó el método de evaluación de impacto cuasiexperimental de diferencias en diferencias, aunado con metodologías de emparejamientos, como el Propensity Score Matching a fin de contrarrestar el sesgo de selección. Los resultados que obtuvo muestran que el impacto en los ingresos del acceso a internet fijo en el hogar es, en promedio, de S/. 298.5 mensuales por año, siendo de S/. 275.8 mensuales a nivel urbano, y de S/. 390.9 mensuales a nivel rural. En tanto, el impacto del uso de internet a nivel rural asciende a S/. 212.1 mensuales por año, mientras que a nivel urbano este impacto no es significativo. En consecuencia, concluyó que el acceso a internet



fijo en el hogar genera un mayor impacto que el acceso a internet a través de diversos dispositivos y en diferentes lugares, incluso fuera del hogar (como en las cabinas públicas, centros de estudios o de labores).

De Los Ríos (2010), analizó el impacto del uso de internet sobre el bienestar de los hogares peruanos. Para desarrollar este estudio, utilizó un panel de viviendas de la Encuestas Nacional de Hogares, para los años 2007 y 2009. El método que aplicó para estimar el impacto del uso de internet sobre el cambio en los ingresos fue los Mínimos Cuadrados Ordinarios (MCO) con regresión lineal, mientras que, para determinar la probabilidad de encontrar trabajo con el uso del servicio de internet, utilizó el método de la Máxima Verosimilitud con regresión logística. Los resultados que obtuvo indican que el impacto de la adopción de internet en el cambio en los ingresos laborales es de S/. 428, equivalentes a 6% del ingreso promedio de los trabajadores en el año 2007. Además, los resultados muestran que no hay una relación estadísticamente significativa entre la probabilidad de conseguir trabajo y el acceso a internet.

Flores et al. (2020), analizaron las tecnologías de información en el Perú con la finalidad de determinar el acceso a internet y brecha digital existente. La metodología de este estudio fue de tipo cualitativo y documental. Dentro de los principales resultados obtuvieron que los adolescentes y jóvenes son los que tienen mayor acceso a internet (80.5%) y que el dispositivo móvil fue el recurso más empleado para tal fin (81.0%). En ese sentido, concluyó que la relación entre la edad y el uso de internet se mantiene aun cuando puedan variar el recurso TIC para acceder a dicho servicio. La relación entre el acceso y uso de internet respecto a nivel educativo se mantiene cuando se comparan variables tales como el género



y el nivel de estudios; de lo cual se desprende que particularmente en el caso peruano, un graduado universitario es por lo menos dos veces más afín a emplear internet en relación con una persona que no ha logrado completar sus estudios primarios.

Ruiz & Ortiz (2014), analizaron los impactos que tiene el acceso y uso de los servicios de internet en los hogares peruanos considerando información desagregada, en base a la información extraída de la Encuesta Nacional de Hogares (ENAHO) en su versión de datos de panel para el período 2007-2009. La metodología que aplicó es el método de Diferencias en Diferencias (DID) y una variante que permite controlar la heterogeneidad en las condiciones iniciales (PSM-DID). Los resultados que obtuvieron muestran que el acceso a internet impacta de manera positiva al ingreso y al gasto total de los hogares. El impacto en el gasto se canaliza a través del mayor gasto en alimentos y en entretenimiento. Asimismo, el acceso al internet impacta de manera positiva en las dimensiones de educación y trabajo. En efecto, los autores evidenciaron impactos positivos en el número de miembros del hogar con educación superior y con empleo adecuado respecto a la PEA.

Roque (2019), evaluó el impacto del acceso y uso de los servicios de internet en el ingreso de los hogares peruanos en la zona urbana y rural del Perú en los años 2015 –2017, en base a la Encuesta Nacional de Hogares (ENAHO). Para lo cual empleó el modelo de Diferencias en Diferencias estimado por el índice de Propensión o Propensity Score en el caso del acceso al servicio de internet, y un modelo de Lineal General estimado por Mínimos Cuadrados Ordinarios (MCO) en el caso del uso al servicio de internet. Los resultados que



obtuvo respecto al impacto en los ingresos económicos de los hogares peruanos de las familias que cuentan con acceso al servicio de internet, en comparación a aquellas familias que no cuenta, asciende a 580 soles mensuales adicionales por hogar, mientras que, respecto al impacto del gasto mensual de estas familias aumenta en 456 soles. Por otro lado, respecto al impacto del uso de servicio de internet respecto a aquellas familias que no hacen uso del servicio de internet, evidenció que 1 sol destinado al consumo de este servicio incrementa el nivel de gasto de los hogares en un 0.37% al mes.

Tovar (2021), evaluó el impacto del acceso al servicio de internet en los ingresos laborales per cápita de los hogares de la región central durante el año 2019, utilizando la Encuesta Residencial de Telecomunicaciones (ERESTEL, 2019) y haciendo el uso de la metodología de evaluación de impacto del Propensity Score Matching (PSM) – Nearest Neighbors. Sus resultados de la evaluación evidencian que un hogar con acceso al servicio de internet genera ingresos laborales per cápita superiores en S/. 272.80 comparado con los hogares que no cuentan con dicho servicio. De acuerdo a los resultados que obtuvo de manera desagregada concluye que el impacto es mucho mayor en hogares donde el jefe de hogar es de sexo femenino, teniendo, en promedio, ingresos laborales per cápita superiores en S/. 119.56, comparado a los hogares con jefes de hogar de sexo masculino, no obstante, la significancia del resultado para el caso femenino no resulta ser tan contundente. Por otra parte, desagregando los efectos según área de residencia, mostró que los hogares que residen en zonas urbanas tienen ingresos laborales per cápita superiores en S/. 316.84, comparado con hogares rurales.



### 2.1.3. Antecedentes locales

Humpiri (2022), analizó los principales factores socioeconómicos que influyen en el acceso a internet en el Perú durante el periodo 2016 – 2019, en base a la Encuesta Nacional de Hogares (ENAHOG). Para tal fin, estimó el modelo Logit con datos de panel, considerando en primera instancia el modelo de efectos fijos como el más adecuado. Los resultados que obtuvo indican que, si se incrementa en un nuevo sol el ingreso del hogar, la probabilidad de que el hogar tenga acceso a internet aumenta en 0.00044; si la edad del jefe del hogar aumenta en 1 año adicional, la probabilidad de que el hogar tenga acceso a internet aumenta en 0.01247; el ser pobre disminuye la probabilidad de que el hogar tenga acceso a internet en 0.74068; si la educación del jefe de hogar aumenta en 1 año adicional, la probabilidad de que el hogar tenga acceso se incrementa en 0.03349; y si el hogar cuenta con electricidad y mayor número de miembros, la probabilidad de que el hogar tenga acceso a internet aumenta en 0.07274 y 0.22798, respectivamente, sin embargo, estas variables no fueron significativas. De acuerdo a los resultados concluyó que los principales determinantes socioeconómicos que influyen al acceso a internet son: el ingreso del hogar, los años de educación del jefe de hogar, la edad del jefe de hogar (de forma positiva) y la pobreza (de forma negativa).

Arias (2023), determinó la relación del uso del internet y el aprendizaje en los estudiantes de la Escuela Profesional de Educación Primaria, mediante la encuesta por fichas para la primera variable y la información en semestres académicos para la segunda variable. La metodología que utilizó fue cuantitativa de tipo descriptivo correlacional, ya que se caracteriza por la relación que existe



entre ambas variables, usando de esta manera la prueba R de Pearson. La primera variable lo conformó por tres dimensiones: frecuencia de uso del internet, el internet en los trabajos académicos y el internet en la ampliación de conocimientos, mientras que para la segunda variable referente al aprendizaje lo conformó por una sola dimensión; el logro de aprendizaje de los estudiantes en sus promedios del semestre académico. Para el análisis de los datos utilizó los programas estadísticos Excel y SPSS. Los resultados que obtuvo evidencian que, entre el uso de internet y el aprendizaje, se encuentra en el intervalo positivo  $r \leq 1$ , con un valor de 0.860 próximo a 1. Con este resultado concluyó que existe una correlación alta entre el uso del internet y el aprendizaje en los estudiantes de la Escuela Profesional de Educación Primaria de la UNA Puno.

Fora (2023), determinó el grado de relación entre el uso de las TIC y el aprendizaje del idioma inglés en los estudiantes del IX Ciclo de curso de inglés de la Facultad de Ingeniería Economía de la Universidad Nacional del Altiplano en el año 2022. Para tan fin, se basó en una investigación de nivel descriptivo – correlacional, de método hipotético deductivo, de diseño no experimental y de corte transversal, considerando la participación de 54 estudiantes del curso de inglés, a los cuales aplicó instrumentos construidos en base a las variaciones, dimensiones e indicadores, mediante la aplicación del estadígrafo Rho de Spearman. Los resultados que obtuvo muestran que las TIC inciden en la enseñanza del idioma inglés en estudiantes del IX Ciclo de curso de inglés de la Facultad de Ingeniería Economía de la Universidad Nacional del Altiplano en el año 2022, debido a que el Rho de Spearman fue del 0.772, indicando una correlación positiva alta. En consecuencia, concluyó que la incorporación de las TIC en el proceso de enseñanza y aprendizaje en las universidades, es de mucha



utilidad para los docentes de educación superior, en cuanto a la enseñanza del idioma inglés, ya que tiene repercusiones sociales y económicos en el país y en el mundo.

## 2.2. MARCO TEÓRICO

### 2.2.1. Modelo de aceptación tecnológica (TAM)

El modelo de aceptación de tecnología, conocido por sus siglas en inglés TAM (*technology acceptance model*) es una teoría de sistemas de información que modela cómo los usuarios llegan a aceptar y utilizar una tecnología (Davis, 1989). El modelo sugiere que cuando los usuarios se enfrentan con una tecnología nueva, existen un conjunto de factores que influyen en su decisión sobre cómo y cuándo lo utilizarán (Davis, 1989).

Este modelo se utiliza para predecir el uso de las TIC, basándose en dos características principales:

- Utilidad percibida (Perceived Usefulness).
- Facilidad de uso percibida (Perceived Ease of Use).

La utilidad percibida (PU) se refiere al grado en que una persona cree que usando un sistema en particular mejorará su desempeño en el trabajo, y la facilidad de uso percibida (PEOU) señala hasta qué grado una persona cree que usando un sistema en particular realizará menos esfuerzo para desempeñar sus tareas (Yong et al., 2010).

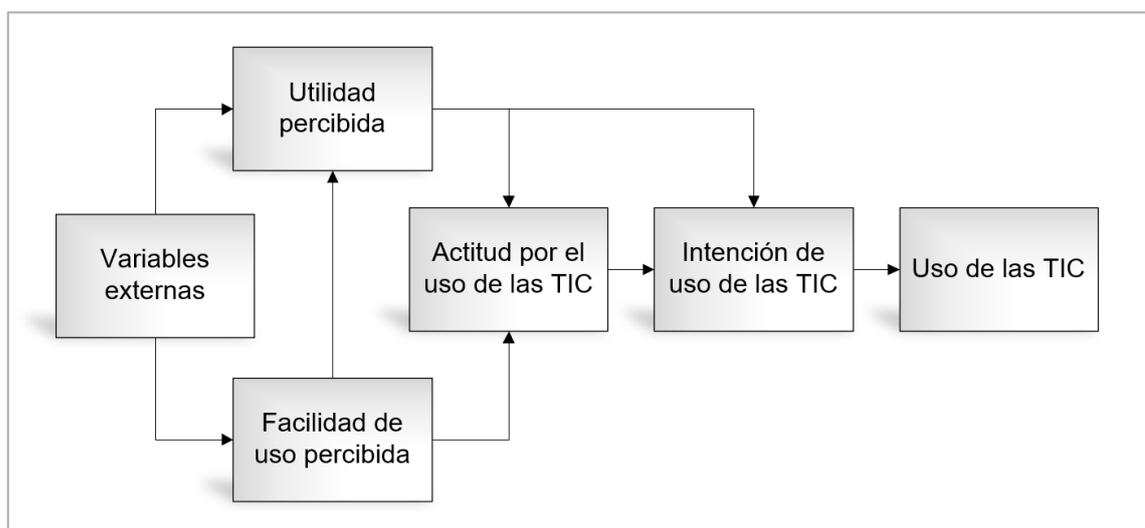
El propósito del TAM es explicar las causas de aceptación de las tecnologías por los usuarios, por lo tanto, este modelo propone que las

percepciones de un individuo en cuanto a la utilidad y la facilidad de uso percibidas de un sistema de información son concluyentes para determinar su intención de usar un sistema (Davis,1989, como se citó en Yong et al., 2010).

De acuerdo con este modelo, existen variables externas que influyen de manera directa en la PU y la PEOU (Yong et al., 2010). Por medio de esta influencia directa en ambas percepciones, las variables externas participan de forma indirecta en la actitud hacia el uso, la intención conductual para usar y la conducta de uso real (Yong et al., 2010). La PEOU tiene un efecto causal en la PU, además del efecto significativo de esta variable en la actitud del usuario (un sentimiento en favor o en contra) hacia el uso del sistema (Yong et al., 2010). El modelo TAM se puede observar en la Figura N° 1.

**Figura 1**

*Modelo de aceptación tecnológica (TAM)*



Fuente: Davis et al. (1989)

El propósito primario del TAM es indagar las consecuencias de los factores externos en cuanto a la utilidad y la facilidad de uso percibidas, para adelantar o predecir el uso de las TIC. Si bien el modelo TAM ayuda a conocer si



una tecnología será utilizada de manera óptima, es necesario identificar las variables externas que inciden en ella, como las causantes de influir de manera directa en la utilidad y la facilidad de uso percibidas por los usuarios de las TIC y determinar la relación de dichas variables con el resultado de su uso (Yong et al., 2010).

### **2.2.2. Teoría de los recursos y de la apropiación de nuevas tecnologías**

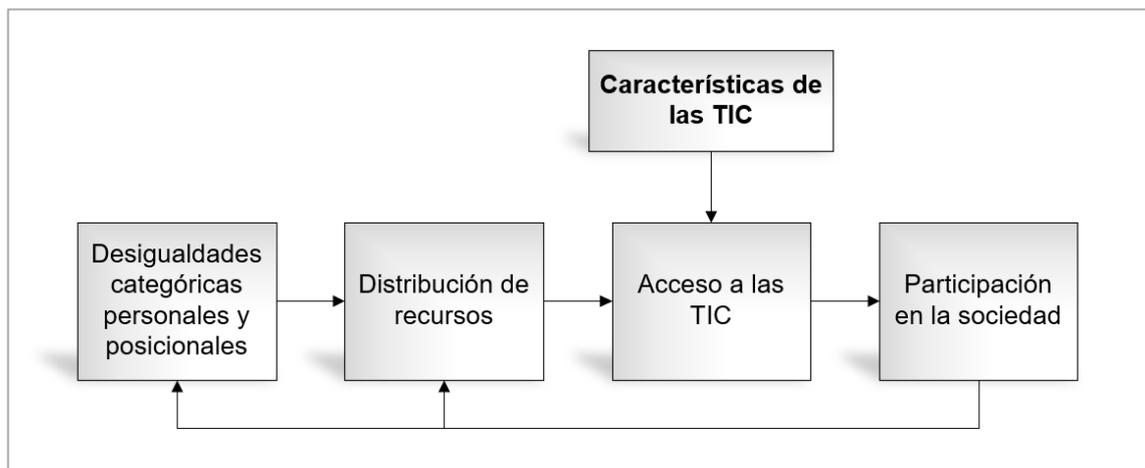
Van Dijk (2012) ha desarrollado la teoría de recursos y apropiación de la difusión, aceptación y adopción de nuevas tecnologías. Los siguientes cuatro son los conceptos centrales de esta teoría:

- Una serie de desigualdades categóricas personales y posicionales en la sociedad.
- La distribución de recursos relevantes para este tipo de desigualdad.
- Diversos tipos de acceso a las TIC.
- Varios ámbitos de participación en la sociedad.

El primero y segundo se consideran las causas, el tercero es el fenómeno a explicar y el cuarto, la posible consecuencia de todo el proceso (Van Dijk, 2012). Al ser parte de un proceso, el cuarto se retroalimenta del primero y segundo, ya que una mayor o menor participación en varios campos de la sociedad cambiará las relaciones de desigualdades categóricas y la distribución de recursos en la sociedad (Van Dijk, 2012). Finalmente, se debe añadir como factor colateral un quinto estado de cosas que determina el tipo de desigualdad a explicar: las características especiales de las tecnologías de la información y la comunicación (Van Dijk, 2012). De esta manera, se puede representar un modelo dinámico de esta teoría. (Ver Figura N° 2).

## Figura 2

### *Modelo causal de recursos y teoría de la apropiación*



Fuente: Van Dijk (2012)

Las características de las TIC como tecnología son factores secundarios en la Figura N° 2. Cuando se considera que una tecnología es compleja, costosa, multifacética (multimedia) y genera problemas de accesibilidad y usabilidad, esto aumentará los problemas de acceso en general. Los dispositivos informáticos simplemente no son iguales a, por ejemplo, los televisores. En las primeras décadas de existencia de las TIC las características mencionadas estaban generalizadas en la oferta de esta tecnología. En la última década se han logrado avances considerables para hacer que el hardware y el software en cuestión sean más accesibles y utilizables para una mayor parte de la población. Es comprensible que esto haya reducido las brechas de habilidades y uso digitales.

El último factor en la Figura N° 2 es lo que está en juego o la preocupación por la brecha digital. Las consecuencias del acceso desigual de todo tipo son una mayor o menor participación en varios campos de la sociedad: económico (como el empleo), social (por ejemplo, contactos sociales), político (votación y otros tipos de participación política), cultural (participación en la cibercultura), espacial



(poder llevar una vida móvil) e institucional (realizar los derechos de ciudadanía) (Van Dijk, 2012).

En ese sentido, el argumento central se puede resumir en las siguientes afirmaciones:

- Las desigualdades categóricas en la sociedad producen una distribución desigual de los recursos.
- Una distribución desigual de los recursos provoca un acceso desigual a las tecnologías digitales.
- El acceso desigual a las tecnologías digitales también depende de las características de estas tecnologías.
- El acceso desigual a las tecnologías digitales genera una participación desigual en la sociedad.
- La participación desigual en la sociedad refuerza las desigualdades categóricas y la distribución desigual de los recursos (Van Dijk, 2012).

Las siguientes desigualdades categóricas personales se pueden observar con frecuencia en la investigación sobre la brecha digital:

- Edad (joven/mayor).
- Género (masculino/femenino).
- Raza/etnicidad (mayoría/minoría).
- Inteligencia (alta/baja).
- Personalidad (extravertida/introvertida, segura de sí misma/no segura de sí misma).
- Salud (capacitados/discapacitados) (Van Dijk, 2012).



Lo mismo ocurre con las siguientes desigualdades categóricas posicionales:

- Posición laboral (empresarios/trabajadores, directivos/empleados, ocupados/desempleados).
- Educación (alta/baja).
- Hogar (familia/persona soltera).
- Nación (desarrollada/en desarrollo) (Van Dijk, 2012).

En la mayoría de las observaciones empíricas, la primera de estas categorías relacionales tiene más acceso que la segunda.

Los siguientes recursos aparecen con frecuencia en la investigación sobre la brecha digital, a veces bajo otras etiquetas, como capital económico, social y cultural:

- Temporal (tener tiempo para utilizar medios digitales).
- Material (posesión e ingresos).
- Mental (capacidad técnica, motivación).
- Social (tener una red social para ayudar en el uso de los medios digitales).
- Cultural (status y gusto por estar en el mundo de los medios digitales) (Van Dijk, 2012).

La parte central del modelo es una serie de tipos de acceso consecutivos. Aquí el concepto multifacético de acceso se refina y se concibe como el proceso total de apropiación de una nueva tecnología. Esto es en parte responsable del nombre de la teoría de recursos y teoría de la apropiación. Para apropiarse de una nueva tecnología, primero hay que estar motivado para utilizarla. Cuando se

desarrolla suficiente motivación, uno debería poder adquirir acceso físico a una computadora, internet u otro medio digital. Además, se necesitan recursos materiales para seguir utilizando la tecnología que consisten en equipos periféricos, software, tinta, papel, suscripciones, etc. Tener acceso físico y material no conduce automáticamente a la apropiación de la tecnología, ya que primero hay que desarrollar varias habilidades para utilizar el medio en cuestión. Cuanto más se desarrollen estas habilidades, más apropiado se podrá hacer el uso de la tecnología en varias aplicaciones. El uso, entre otras cosas, se refiere a la frecuencia de uso y al número y diversidad de aplicaciones. Por lo tanto, este proceso se representa en la Figura N° 3.

### Figura 3

#### *Etapas de acceso en la apropiación de la tecnología digital*



Fuente: Van Dijk (2012)



### **2.2.2.1. Motivación**

La motivación es la primera etapa en el proceso de la apropiación de una nueva tecnología. Las personas a las que no les gustan las computadoras u otros medios digitales no intentarán comprar una ni obtener una conexión en particular, a menos que se vean obligadas a hacerlo (Van Dijk & Van Deursen, 2014).

### **2.2.2.2. Acceso físico y material**

La segunda etapa de apropiación, la adquisición de acceso físico a los medios digitales, ha dominado por completo la opinión pública y las perspectivas políticas en las últimas dos décadas. El acceso físico no es igual al acceso material, ya que este último incluye todos los costos del uso de computadoras, conexiones, equipos periféricos, software y servicios (Van Dijk & Van Deursen, 2014). Estos costos divergen en muchos sentidos, y las personas con acceso físico tienen gastos bastante diferentes en computadora, internet y otros medios digitales. Aunque la brecha de acceso físico se cerrará en el largo plazo, las brechas de acceso material permanecerán y tal vez se vuelvan aún más prominentes (Van Dijk & Van Deursen, 2014). La innovación de las tecnologías de la información y la comunicación (TIC) no se frena y continuamente se inventan nuevos hardware y software más o menos costosos (Van Dijk & Van Deursen, 2014). La innovación en los servicios también genera mayores o menores gastos para quienes tienen diferentes necesidades e ingresos (Van Dijk & Van Deursen, 2014).



### 2.2.2.3. Habilidades digitales

Después de haber adquirido la motivación para utilizar computadoras y algún tipo de acceso físico a ellas, se debe aprender a manejar el hardware y el software (Van Dijk, 2012). Este problema se enmarca con términos como “alfabetización informática, informacional o multimedia” y “habilidades informáticas” o “cápita de información”. En ese contexto, se ha definido el concepto de habilidades digitales en seis tipos (Van Dijk, 2012):

- **Habilidades relacionado con el medio**

- **Habilidades operativas:** Acciones necesarias para operar un medio digital (conocimiento de botones).
- **Habilidades Formales:** Manejo de las estructuras formales del medio (navegación).

- **Habilidades relacionado con el contenido**

- **Habilidades Informativas:** Buscar, seleccionar y evaluar información en medios digitales, por ejemplo, motores de búsqueda.
- **Habilidades de comunicación:** Enviar correos, contactar, crear identidades en línea, llamar la atención y dar opiniones.
- **Habilidades de creación de contenido:** Hacer contribuciones a internet con un plan o diseño particular.
- **Habilidades estratégicas:** Utilizar el medio digital como medio para alcanzar determinados objetivos profesionales y personales.



#### 2.2.2.4. Uso de las TIC

Evidentemente, el propósito del proceso total de apropiación es el uso. Tener suficiente motivación, acceso físico y habilidades para aplicar los medios digitales son condiciones necesarias, pero no suficientes para el uso real (Van Dijk, 2012). El uso tiene sus propios fundamentos o determinantes, ya que como factor dependiente se puede medir al menos de cuatro maneras (Van Dijk, 2012):

- Tiempo y frecuencia de uso.
- Número y diversidad de aplicaciones de uso.
- Uso de banda ancha o banda estrecha.
- Uso más o menos activo o creativo (Van Dijk, 2012).

Además, los patrones de uso consisten en la frecuencia y la duración del día en que se utiliza el medio digital, el número y variedad de aplicaciones, los tipos de aplicaciones utilizadas (por ejemplo, información, comunicación, comercio, trabajo, entretenimiento y educación) y el tipo de uso (productivo y generado por el usuario o consuntivo). Todas estas variaciones están correlacionadas con los datos demográficos que se examinan con frecuencia en las investigaciones sobre la brecha digital, es decir, la edad, el género, el nivel educativo, la ocupación, la composición del hogar y el origen étnico (Van Dijk & Van Deursen, 2014).



### **2.2.3. Teorías del empleo**

#### **2.2.3.1. Teoría del empleo marxista**

La teoría marxista del empleo parte de la oferta para explicar las causas del empleo, siendo este producto del proceso de acumulación capitalista. La demanda de trabajo depende del capital variable, por lo que la disminución proporcional en su inversión provoca un exceso de mano de obra, formándose el ejército industrial de reserva. (Colás, 2007)

#### **2.2.3.2. La teoría del empleo keynesiana**

La teoría keynesiana plantea que el exceso de la demanda global es la causante del desempleo y que el mercado, en épocas de crisis, no es capaz de regular automáticamente la economía, principalmente para mantener el nivel de empleo. En el corto plazo, son las rigideces de los salarios nominales las que impiden el ajuste del mercado de trabajo. (Colás, 2007)

#### **2.2.3.3. La teoría del empleo neoclásica**

La teoría neoclásica plantea que la intervención del Estado y de los sindicatos provocan el aumento de los salarios reales por encima del nivel de equilibrio y por tanto, el desempleo. De no existir dicha intervención, la economía trabaja a un nivel de pleno empleo, donde la demanda de trabajo es igual a la oferta de trabajo. (Colás, 2007)

### **2.2.4. Tipos de desempleo**

Los neoclásicos definen tres tipos de desempleo:



- **Desempleo friccional:** Es el desempleo que se debe principalmente a las bajas voluntarias, a los cambios de trabajo y a las personas que entran por primera vez en la población activa o que retornan a ella. Incluso aunque una economía tuviera pleno empleo, siempre habría rotación, ya que los individuos buscan trabajo cuando terminan sus estudios o se trasladan de una ciudad a otra, etc. Algunos lo consideran como un tipo de desempleo estructural (Larraín & Sachs, 1993, como se citó en Colás, 2007).
- **Desempleo estructural:** Es el desempleo que forma parte de la tasa natural de desempleo de un país este tipo de desempleo tiene muchas características comunes con el desempleo friccional, pero se diferencian en que es más prolongado. Por lo tanto, puede tener grandes costos para los desempleados y suponer una pérdida considerable de producción necesarias para ocupar los puestos de trabajo existentes y las que poseen los demandantes de empleo; o un desajuste geográfico entre el lugar donde se encuentran las vacantes y el lugar donde se encuentra los demandantes de empleo (Larraín & Sachs, 1993, como se citó en Colás, 2007).
- **Desempleo cíclico:** Existe cuando hay un nivel insuficiente de demanda agregada, lo que obliga a las empresas a suspender temporalmente el empleo a los trabajadores o despedirlos. Cuando disminuyen el gasto y la producción totales, el desempleo aumenta en casi todas partes. No existe como consecuencia de las diferencias entre las tasas de inflación esperada y efectiva. La distinción entre desempleo cíclico y otros, ayuda a los economistas a diagnosticar la salud general del mercado de trabajo. Existe desempleo cíclico cuando disminuye el empleo a consecuencia de la oferta



y demanda agregada, como subproducto de las recesiones y las depresiones (Larraín & Sachs, 1993, como se citó en Colás, 2007).

### **2.2.5. Teoría del bienestar**

El concepto de bienestar es un concepto difícil de precisar, debido a los aspectos subjetivos que afectan al bienestar individual y se trasladan al bienestar colectivo o social, además, nadie sostiene la total equivalencia entre riqueza y bienestar, por lo que no se debe utilizar el PNB como medida de bienestar de un país y se acepta que el bienestar engloba otros aspectos (Reyes & Franklin, 2014).

Desde el campo de la economía el estudio del bienestar ha sido una preocupación constante, sobre la cual diferentes escuelas y pensadores han realizado diversos aportes que sirven como referentes para la formulación de nuevas teorías de bienestar, porque la aplicación de la teoría del bienestar en formar racional conduce a minimizar los niveles de desigualdad social y a establecer un justa medida de distribución de los escasos recursos con que cuenta la sociedad (Reyes & Franklin, 2014).

Es posible medir el bienestar social a partir del enfoque económico, de las funciones de utilidad y utilizando indicadores sociales, aunque cada uno de estos enfoques, presentan dificultades y debilidades en su medición, por lo tanto, el concepto de utilidad visto a la luz del bienestar es un concepto relativo, por cuanto depende de la valoración que cada individuo le asigne y cualquier cambio, que no perjudique a un individuo y que proporcione a alguna persona calidad de vida, debe considerarse como una mejoría para el bienestar de la sociedad (Reyes & Franklin, 2014).



El equilibrio general competitivo constituye un óptimo paretiano, pero no todo equilibrio paretiano es un equilibrio general competitivo, por lo que este criterio plantea una disyuntiva entre eficiencia y equidad, ya que si bien resuelve el óptimo individual no resuelve el problema del óptimo social, donde no sólo es relevante la asignación de los recursos, sino también la distribución equitativa de la renta (Reyes & Franklin, 2014).

#### **2.2.6. Economía digital**

La economía digital está constituida por la infraestructura de telecomunicaciones, las industrias TIC (software, hardware y servicios TIC) y la red de actividades económicas y sociales facilitadas por internet, la computación en la nube y las redes móviles, las sociales y de sensores remotos (CEPAL, 2013). La economía digital es un facilitador cuyo desarrollo y despliegue se produce en un ecosistema caracterizado por la creciente y acelerada convergencia entre diversas tecnologías, que se concreta en redes de comunicación (redes y servicios, redes fijas-móviles), equipos de hardware (móviles multimedia 3G y 4G), servicios de procesamiento (computación en la nube) y tecnologías web (Web 2.0) (CEPAL, 2013).

La economía digital consta de tres componentes principales:

- La infraestructura de redes de banda ancha: La infraestructura de redes de banda ancha es el primer componente de la economía digital, por lo que los elementos básicos de esa infraestructura son la conectividad nacional e internacional, las redes de acceso local, los puntos de acceso público y la asequibilidad (CEPAL, 2013).



- La industria de aplicaciones TIC: El segundo componente esencial para generar servicios y aplicaciones para los usuarios (individuos, empresas y gobierno) es la industria de hardware, software y aplicaciones TIC que incluye también servicios facilitados por estas tecnologías (CEPAL, 2013).
- Los usuarios finales: El tercer componente son los usuarios finales (individuos, empresas y gobierno) que definen el grado de absorción de las aplicaciones digitales mediante su demanda por servicios y aplicaciones (CEPAL, 2013).

### **2.2.7. Teletrabajo**

El teletrabajo se caracteriza por el desempeño subordinado de labores sin la presencia física del trabajador, denominado “teletrabajador”, en la empresa con la que mantiene vínculo laboral, a través de medios informáticos, de telecomunicaciones y análogos, mediante los cuales se ejercen a su vez el control y la supervisión de las labores (Ley N° 30036: Ley que regula el Teletrabajo. Art. 2, 2013). Son elementos que coadyuvan a tipificar el carácter subordinado de esta modalidad de trabajo la provisión por el empleador de los medios físicos y métodos informáticos, la dependencia tecnológica y la propiedad de los resultados, entre otros (Ley N° 30036: Ley que regula el Teletrabajo. Art. 2, 2013).

El teletrabajador requiere del conocimiento y uso intensivo de las TIC en el desarrollo de sus actividades, sin embargo, debe considerarse otro aspecto importante, como lo es la distancia, es decir, la persona trabaja fuera o distante de las oficinas del negocio o empresa para la cual labora (Osio, 2010).



### **2.2.8. Impacto de las TIC en el empleo**

La tecnología ha tenido un gran impacto en el empleo, por un lado, ha reducido la necesidad de la mano de obra gracias a la automatización de muchas tareas, pero también, ha creado nuevos empleos en muchos sectores, además de una mejora en la eficiencia y la productividad de las industrias (Mantecón, 2023).

“Para estudiar el efecto que esto ha tenido en el empleo, se hace distinción entre los dos tipos de trabajos que han traído consigo las nuevas tecnologías, el trabajo autoprogramable y el trabajo genérico” (Mantecón, 2023). En cuanto al trabajo autoprogramable, es aquel que posee una persona capaz de adaptarse a los distintos retos que la tecnología presenta, sin embargo, el trabajo genérico hace referencia a aquellos que son fácilmente reemplazables, ya que sus tareas son rutinarias y se dedican únicamente a recibir instrucciones y ejecutar órdenes (Mantecón, 2023). “Todos los trabajos incluidos en este último punto son los que las máquinas están reemplazando por su bajo requerimiento de conocimientos técnicos” (Mantecón, 2023).

### **2.2.9. Efectos de la pandemia de COVID-19 sobre la adopción de las TIC en el Perú**

De manera previa a los acontecimientos desencadenados por la pandemia de la COVID-19, las tecnologías de la sociedad de la información se venían implementando en el Perú de manera gradual pero sostenida, sin embargo, a pesar de este ritmo de crecimiento estable, nuestro país era uno de los menos interconectado de la región (Huarag, 2020).



La pandemia de la COVID-19 ha sido un verdadero desastre sin paliativos para el país, por lo que la existencia de una pronunciada brecha digital en el Perú ha privado a muchos ciudadanos de herramientas esenciales para enfrentar esta crisis mediante el aislamiento social (Huarag, 2020). Los beneficios potenciales de las tecnologías de la información y comunicación (TIC) solo pueden ser empleados por quienes tienen acceso a internet, pero deja indefensos a quienes no tienen acceso a la infraestructura o a los servicios de la sociedad de la información (Huarag, 2020).

El bajo índice de penetración de internet en los hogares impide la adopción de estrategias potencialmente más exitosas para afrontar la pandemia de la COVID-19 mediante la utilización de las TIC, por lo tanto, no es fácil, con índices tan bajos de digitalización, reemplazar las actividades cotidianas que requieran de presencia física por actividades digitales (Huarag, 2020).

#### **2.2.10. Impacto del COVID-19 sobre el empleo en el Perú**

El surgimiento de la pandemia en el 2020 generó una respuesta de política gubernamental caracterizada por la implementación de cuarentenas obligatorias y restricciones de movilidad, las cuales afectaron a la dinámica de los mercados laborales en todo el mundo, en efecto, las repercusiones de la COVID-19 sobre el empleo han sido diferenciadas no solo entre países sino también al interior de los mismos (Lavado, 2021).

Específicamente para el caso peruano, un análisis desagregado a lo ocurrido durante el segundo trimestre del 2020, muestra que la caída de la población ocupada – una de las más severas en comparación a otras economías de América Latina – fue heterogénea, y afectó en mayor medida a los trabajadores



dependientes e independientes del sector informal de la economía (Lavado, 2021). Por el contrario, el empleo formal registró una menor contracción, lo cual pone en evidencia la rigidez normativa y menor flexibilidad a la cual se encuentra expuesta este segmento del mercado de trabajo para ajustarse marginalmente incluso en periodos de crisis (Lavado, 2021).

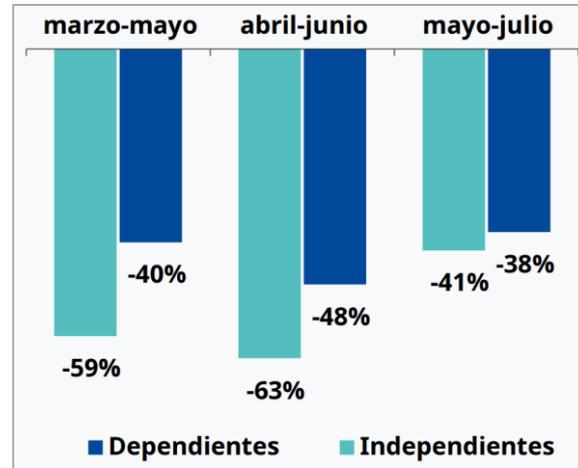
#### **2.2.10.1. Impacto según tipo de trabajador**

Tanto trabajadores dependientes como independientes fueron afectados por la pandemia, aunque el mayor impacto se registró en el segundo grupo, en ese sentido, los datos de la Encuesta Permanente de Empleo (EPE) para Lima Metropolitana muestran que el empleo entre los independientes se redujo en términos interanuales en 59% en el trimestre marzo-mayo y en 63% en abril-junio, en cambio, la reducción del empleo entre los asalariados fue de 40% y 48%, respectivamente (Organización Internacional del Trabajo [OIT], 2021a).

Más recientemente, en el trimestre mayo-junio, la caída del empleo entre los independientes (-41%) y dependientes (-38%) presentó una mayor similitud, debido a que el levantamiento de las medidas restrictivas facilitó que los trabajadores independientes pudieran realizar actividades que generan ingresos (OIT, 2021a).

**Figura 4**

*Reducción del empleo según tipo de trabajador, 2020 vs. 2019, a nivel de Lima Metropolitana*



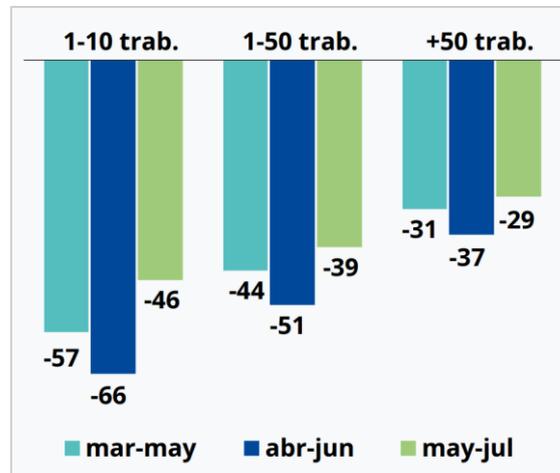
Fuente: Instituto Nacional de Estadística e Informática

### 2.2.10.2. Impacto según el tamaño de la empresa

Los datos de la EPE muestran que los trabajadores de las empresas más pequeñas han sido los más perjudicados, por ejemplo, en el trimestre de mayor impacto (abril-junio), el empleo en las empresas de menos de diez trabajadores se redujo en 66%, porcentaje superior al de aquellas con once a cincuenta trabajadores (-51%) y aquellas de más de cincuenta trabajadores (-37%) (OIT, 2021a).

**Figura 5**

*Reducción del empleo según tamaño de empresa, 2020 vs. 2019, a nivel de Lima Metropolitana*



Fuente: Instituto Nacional de Estadística e Informática

### 2.2.10.3. Impacto según actividad económica

El impacto en el empleo según actividad económica está fuertemente relacionado con la paralización de las actividades en cada sector, en general, las actividades primarias fueron las menos afectadas, por ejemplo, las operaciones del sector agropecuario no fueron restringidas ya que son consideradas esenciales, por lo que durante el segundo trimestre del año, su producción se incrementó en 1%, en contraste con la mayoría de actividades económicas que mostraron resultados negativos, en consecuencia, el empleo del sector en el mismo periodo de se incrementó en 22.6% (OIT, 2021a).

“Por su parte, otras actividades primarias como la minería y la pesca sí experimentaron un mayor impacto, principalmente durante abril, aunque, debido al reinicio de las actividades económicas en mayo, lograron reactivar sus operaciones” (OIT, 2021a). En particular, con el



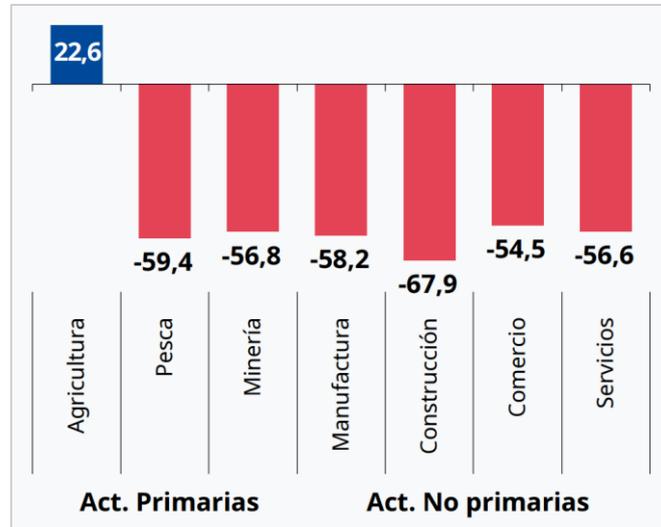
inicio de la primera temporada pesquera del año, la producción pesquera, que tiene un fuerte componente estacional, se incrementó en 48% en junio, por su parte, la actividad minera muestra cada vez menores caídas interanuales y se encuentra cerca de alcanzar niveles de producción previos a la emergencia, de este modo, mientras que en abril y mayo el VAB sectorial se redujo por encima del 40%, en junio, la reducción interanual fue de 14.4% (OIT, 2021a).

En este caso, el empleo en estos sectores sí se contrajo considerablemente durante el segundo trimestre del año: 59.4% en pesca y 56.8% en minería, no obstante, se espera que los próximos indicadores disponibles muestren una recuperación de los puestos de trabajo, gracias a la reanudación de sus actividades productivas (OIT, 2021a).

Entre las actividades no primarias, el impacto ha sido mayor y más persistente, ya que, en el país, durante el segundo trimestre, la mayor reducción del empleo en términos porcentuales se registró en el sector construcción (-67,9%), seguido de los sectores manufactura (-58,2%), servicios (-56,6%) y comercio (-54,5%) (OIT, 2021a).

**Figura 6**

*Reducción del empleo según actividad económica, 2020 vs. 2019, a nivel de Lima Metropolitana*



Fuente: Instituto Nacional de Estadística e Informática

#### **2.2.10.4. Impacto en el empleo juvenil**

Los efectos negativos de la pandemia por COVID-19 sobre el empleo se habrían concentrado más en la población joven, por ejemplo, según los datos de la ENAHO, en el segundo trimestre de 2020, la mayor caída del empleo se registró en las personas entre 14 a 24 años de edad, por lo que el empleo para este grupo etario se redujo en 46.7% en términos interanuales, por encima de la reducción registrada para el total de la fuerza laboral (39.6%) (OIT, 2021a).

#### **2.2.11. Modelos econométricos con variables dependientes discretas**

##### **2.2.11.1. Modelos de elección discreta binaria**

Al respecto, Ordaz et al. (2011) afirma lo siguiente:



Los modelos de elección discreta se caracterizan por el hecho de permitir reflejar la elección o toma de decisión por parte de un individuo entre diversas alternativas posibles. Si éstas son solamente dos, se trata de modelos de elección binaria. Por ejemplo, al individuo se le plantea tomar una decisión de entre únicamente dos posibilidades mutuamente excluyentes. La variable endógena  $Y$  de estos modelos adopta dos únicos valores numéricos discretos, normalmente 1 y 0; de modo que, si el sujeto se decanta por la ocurrencia del suceso objeto de estudio entonces  $Y$  toma el valor 1, y 0, si no es así. Estos modelos facilitan la tarea de identificación de las características o factores que inciden en un comportamiento de los individuos diferente ante los procesos de decisión. (Ordaz et al., 2011)

Se puede plantear este tipo de modelos como sigue:

Suponiendo que la variable  $Y$  depende de un conjunto de variables explicativas  $X$ , de manera que:

$$Y_i = F(X_i\beta) + \mu_i, \quad \dots \text{Ecuación 01}$$

donde  $X_i = (X_{1i}, X_{2i}, \dots, X_{ki})$  hace referencia a las observaciones de todas las variables explicativas del modelo, entonces se tiene que:

$$E[Y_i|X_i] = E[F(X_i\beta)] + E[\mu_i] = F(X_i\beta), \quad \dots \text{Ecuación 02}$$

manteniendo el supuesto de que  $E[\mu_i] = 0$ .

Por otra parte, si se calcula la esperanza condicionada de  $Y$  en términos probabilísticos, entonces:

$$\begin{aligned} E[Y_i|X_i] &= \sum_i Y_i \cdot P(Y_i|X_i) = 1 \cdot P(Y_i = 1|X_i) + 0 \cdot P(Y_i = 0|X_i) \\ &= P(Y_i = 1|X_i). \end{aligned}$$

... Ecuación 03

De donde se deduce, pues, que:

$$E[Y_i|X_i] = F(X_i\beta) = P(Y_i = 1|X_i). \quad \dots \text{Ecuación 04}$$

Teniendo en cuenta que la variable  $Y_i$  únicamente puede tomar los valores 1 y 0, el significado del modelo implica que éste asigna cierta probabilidad condicional de que  $Y_i = 1$ , que se denotara por  $P_i$ , es decir:

$$P(Y_i = 1|X_i) = P_i = F(X_i\beta); \quad \dots \text{Ecuación 05}$$

y, en consecuencia:

$$P(Y_i = 0|X_i) = 1 - P_i = 1 - F(X_i\beta). \quad \dots \text{Ecuación 06}$$

El modelo estima, por tanto, la probabilidad para la observación  $i$  de elegir la opción 1, ya que:

$$\hat{Y}_i = F(X_i\hat{\beta}) = \hat{P}_i. \quad \dots \text{Ecuación 07}$$

Dependiendo de la forma funcional concreta que adopte  $F(X_i\beta)$ , se obtienen distintos modelos de elección binaria: modelo de probabilidad lineal y modelo de probabilidad no lineal. Sin embargo, en este caso se analizará solo el modelo de probabilidad no lineal, ya que la presente investigación está basada en ello. (Ordaz et al., 2011)

### 2.2.11.2. Modelo de probabilidad no lineal (Logit y Probit)

Al respecto, Ordaz et al. (2011) indica lo siguiente:

Los modelos Logit y Probit son modelos de elección binaria no lineales muy semejantes. Presentan una serie de propiedades que justifican su utilización:

- Se definen para todo valor de las variables explicativas:  $(-\infty, +\infty)$ .
- Sus funciones de distribución son continuas y toman siempre valores comprendidos entre 0 y 1.
- Cuando  $X_i\beta \rightarrow -\infty \Rightarrow P_i \rightarrow 0$ .
- Cuando  $X_i\beta \rightarrow +\infty \Rightarrow P_i \rightarrow 1$ .
- Incrementan monótonamente respecto a  $X_i\beta$ .

La diferencia básica entre ellos radica en la función de distribución de probabilidad sobre la que se basan,  $F(\cdot)$ . En concreto, el Logit emplea la función logística de distribución:  $\Lambda(\cdot)$  y el Probit, la correspondiente a la distribución normal:  $\Phi(\cdot)$ . Estas funciones de distribución se caracterizan por tener forma de “S”; su punto de inflexión depende de cada una de ellas. (Ordaz et al., 2011)

#### 2.2.11.2.1. Definición del modelo Logit

El modelo Logit es un modelo de respuesta o elección binaria, caracterizado porque la función  $F(X_i\beta)$  se corresponde con la de la distribución logística, de modo que si se denota ésta por  $\Lambda(\cdot)$ , se tiene que (Ordaz et al., 2011):

$$Y_i = \Lambda(X_i\beta) + \mu_i. \quad \dots \text{Ecuación 08}$$

De esta forma:

$$E[Y_i|X_i] = P(Y_i = 1|X_i) = P_i = \Lambda(X_i\beta) = \frac{1}{1 + e^{-X_i\beta}} \quad \dots \text{Ecuación 09}$$

y

$$P(Y_i = 0|X_i) = 1 - P_i = 1 - \frac{1}{1 + e^{-X_i\beta}} = \frac{1}{1 + e^{X_i\beta}}. \quad \dots \text{Ecuación 10}$$

Así pues, la estimación del modelo proporciona la cuantificación de la probabilidad de elegir la opción 1, es decir (Ordaz et al., 2011):

$$\hat{Y}_i = \hat{P}_i = \Lambda(X_i\hat{\beta}) = \frac{1}{1 + e^{-X_i\hat{\beta}}}. \quad \dots \text{Ecuación 11}$$

La interpretación de los parámetros  $\beta_j$  en un modelo Logit no es inmediata, ya que no proporciona, como en el caso de los modelos lineales, el efecto marginal de un cambio unitario en  $X_j$  sobre la probabilidad de elegir la alternativa 1 (la variable endógena  $Y$ ) (Ordaz et al., 2011).

En efecto, dada la no linealidad del modelo, se tiene que (Ordaz et al., 2011):

$$\frac{P_i}{1 - P_i} = \frac{1}{\frac{1 + e^{-X_i\beta}}{1 + e^{X_i\beta}}} = e^{X_i\beta}, \quad \dots \text{Ecuación 12}$$

de donde:

$$X_i\beta = \ln\left(\frac{P_i}{1 - P_i}\right) = L_i. \quad \dots \text{Ecuación 13}$$

$L_i$  recibe el nombre de Logit, y es lo que da nombre al modelo. El Logit representa el logaritmo neperiano de la razón de la probabilidad a favor de la alternativa 1, por tanto (Ordaz et al., 2011):

$$\beta_j = \frac{\partial L_i}{\partial X_{ji}}. \quad \dots \text{Ecuación 14}$$

Es decir, los parámetros miden el cambio en el Logit ocasionado por un cambio unitario en la variable  $X_j$ ; esto es, cuánto varía el logaritmo de la razón de probabilidades a favor de la ocurrencia de la opción 1, ante incrementos unitarios de  $X_j$ , no el efecto marginal de un cambio unitario en  $X_j$  sobre la probabilidad de ocurrencia de la opción 1,  $P_j$  (Ordaz et al., 2011).

Éste último viene dado por la expresión (Ordaz et al., 2011):

$$\frac{\partial P_i}{\partial X_{ji}} = \frac{\partial \Lambda(X_i \beta)}{\partial X_{ji}} = \lambda(X_i \beta) \cdot \beta_j, \quad \dots \text{Ecuación 15}$$

donde  $\lambda(X_i \beta)$  es la función de densidad de la distribución logística.

Esto es, la variación en la probabilidad de la ocurrencia de la opción estudiada ante variaciones unitarias de  $X_j$  viene dada por el producto de  $\beta_j$  por el valor que toma la función de densidad de la distribución logística en la observación  $i$ -ésima (Ordaz et al., 2011).

Este último detalle es importante. La magnitud de la variación de la probabilidad, dado un incremento unitario de la correspondiente variable explicativa, depende de su nivel de partida y, por consiguiente, de los valores de todos y cada uno de los regresores y coeficientes en la



observación donde se estudie. Esto supone, como puede observarse, la superación de la limitación que ofrece el MLP respecto a la consideración de un efecto marginal de las variables explicativas sobre la probabilidad constante para todas las observaciones; supuesto que se consideraba poco ajustado a la realidad (Ordaz et al., 2011).

En el caso de que  $X_j$  sea una variable dummy, el análisis del efecto marginal de ésta sobre la probabilidad se calcula a través de la diferencia de los valores proporcionados por  $E[Y_i|X_{ji} = 1]$  y  $E[Y_i|X_{ji} = 0]$  (Ordaz et al., 2011).

Como el efecto marginal de un cambio unitario de  $X_j$  sobre la probabilidad varía dependiendo de la observación donde se realice, para obtener un valor representativo éste se suele medir en los valores medios de los regresores (Ordaz et al., 2011).

Es importante reseñar que el signo de  $\beta_j$  sí indica el sentido del cambio en la probabilidad, aunque no su cuantía (Ordaz et al., 2011).

La estimación de este modelo se realiza habitualmente a través del método de máxima verosimilitud (MV). Los estimadores MV resultan ser consistentes y asintóticamente eficientes, por lo que este método es recomendable que se emplee para muestras “suficientemente” grandes (Ordaz et al., 2011).

#### **2.2.11.2.2. Definición del modelo Probit**

Otro de los modelos de elección binaria es el modelo Probit. Este modelo se caracteriza porque la función de distribución que utiliza es la

correspondiente a la normal estándar:  $\Phi(\cdot)$ . De este modo, se tiene que (Ordaz et al., 2011):

$$Y_i = \Phi(X_i\beta) + \varepsilon_i. \quad \dots \text{Ecuación 16}$$

Y consiguientemente:

$$E[Y_i|X_i] = P(Y_i = 1|X_i) = P_i = \Phi(X_i\beta) = \int_{-\infty}^{X_i\beta} \phi(s)ds, \quad \dots \text{Ecuación 17}$$

donde  $\phi(s) = \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}}$  es la función de densidad de la distribución normal estándar, con media cero y varianza 1, y  $s$  es una variable “muda” de integración (Ordaz et al., 2011).

Así que, la estimación del modelo ofrece la cuantificación de la probabilidad de elegir la alternativa 1; esto es (Ordaz et al., 2011):

$$\hat{Y}_i = \hat{P}_i = \Phi(X_i\hat{\beta}) = \int_{-\infty}^{X_i\hat{\beta}} \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds. \quad \dots \text{Ecuación 18}$$

En cuanto a la interpretación de los parámetros, en el modelo Probit sucede exactamente igual que en el modelo Logit. Los parámetros estimados no determinan el efecto marginal de variaciones de las variables explicativas  $X_j$  sobre la probabilidad, aunque su signo sí determina el sentido del cambio. El efecto marginal resulta del producto del valor de la función de densidad del modelo (en este caso la normal) en un punto determinado y el parámetro correspondiente (Ordaz et al., 2011):

$$\frac{\partial P_i}{\partial X_{ji}} = \frac{\partial \Phi(X_i\beta)}{\partial X_{ji}} = \phi(X_i\beta) \cdot \beta_j. \quad \dots \text{Ecuación 19}$$



Como se puede observar, la magnitud de las variaciones de la probabilidad depende del nivel donde se observe, por lo que (de forma análoga a como sucede en el modelo Logit) es función de los valores de todas y cada una de las variables explicativas y de sus coeficientes en aquella observación donde se estudie (Ordaz et al., 2011).

Si  $X_j$  fuese una variable ficticia o dummy, el análisis de su efecto marginal sobre la probabilidad, se realiza a través de la diferencia de los valores proporcionados por  $E[Y_i|X_{ji} = 1]$  y  $E[Y_i|X_{ji} = 0]$  (Ordaz et al., 2011).

De nuevo, para obtener un valor representativo de los efectos marginales éstos se suelen medir en los valores medios de los regresores. Al igual que el modelo Logit, el modelo Probit suele estimarse por el método de MV (Ordaz et al., 2011).

- **Contraste de significatividad individual**

La estimación de los parámetros de estos modelos por MV resulta asintóticamente normal, por lo que los contrastes resultan asintóticos. Éstos se realizan de forma similar a como se hace en el modelo de regresión lineal, con la diferencia de que en lugar de emplear estadísticos que siguen la distribución t-Student, en este caso se trabaja con la distribución normal tipificada (Ordaz et al., 2011).

De este modo, para contrastar la hipótesis nula  $H_0: \beta_j = 0$ , se tendrá que (Ordaz et al., 2011):

$$P\left(-N_{\frac{\alpha}{2}} < \frac{\hat{\beta}_j}{ES(\hat{\beta}_j)} < N_{\frac{\alpha}{2}}\right) = 1 - \alpha, \quad \dots \text{Ecuación 20}$$

donde para un nivel de confianza  $1 - \alpha$ ,  $N_{\alpha/2}$  y  $-N_{\alpha/2}$  son los valores críticos (simétricos) de la distribución normal tipificada,  $\hat{\beta}_j$  se refiere a cada uno de los parámetros estimados, y  $ES(\hat{\beta}_j)$  es la estimación del error estándar del parámetro estimado correspondiente.

Como bien se sabe, si la desigualdad se verifica, ello supondrá aceptar la hipótesis nula, con lo que el parámetro  $\beta_j$ , y con ello la variable explicativa  $X_j$ , no serán significativos. Si, por el contrario, no se verificase, querrá decir que la variable en cuestión es relevante en la explicación de la variable dependiente del modelo (Ordaz et al., 2011).

- **Medidas de bondad del ajuste y de contraste de significatividad global de los modelos de elección discreta binaria**

En estos modelos, el habitual coeficiente de determinación  $R^2$  no resulta válido como medida de bondad del ajuste. Por ello, se han desarrollado otras medidas alternativas (Ordaz et al., 2011).

-  $R^2$  de McFadden:

Su expresión es:

$$R^2 \text{ de McFadden} = 1 - \frac{\ln(L)}{\ln(L_R)}, \quad \dots \text{Ecuación 21}$$

donde  $\ln(L_R)$  es el logaritmo neperiano de la función de verosimilitud del modelo restringido, bajo la hipótesis nula  $H_0: \beta_2 = \dots =$

$\beta_k = 0$ , mientras que  $\ln(L)$  es el logaritmo de la función de verosimilitud del modelo original (sin restricciones) (Ordaz et al., 2011).

El valor de este estadístico está comprendido entre 0, que implica nula significatividad del modelo, y 1, que supone un ajuste perfecto, pero para los valores intermedios su significado no es tan claro ni directo como en el caso del  $R^2$  lineal (Ordaz et al., 2011).

- Estadístico LR (LR-statistic) o test de la razón de verosimilitud:

El estadístico LR se define de la forma (Ordaz et al., 2011):

$$LR = -2\ln\left(\frac{L_R}{L}\right), \quad \dots \text{Ecuación 22}$$

donde  $L_R$  es la función de verosimilitud del modelo restringido, bajo la hipótesis nula  $H_0: \beta_2 = \dots = \beta_k = 0$ , y  $L$  es la función de verosimilitud del modelo original (sin restricciones).

Este estadístico sigue una distribución  $\chi^2$  con  $k - 1$  grados de libertad.

El rechazo de  $H_0$  implica que el modelo es, en su conjunto, significativo. Por el contrario, su aceptación implica que no lo es.

- Pseudo  $R^2$  de predicción:

Esta medida indica la proporción de predicciones correctas que realiza el modelo. Concretamente, se define como (Ordaz et al., 2011):

$$\text{Pseudo } R^2 \text{ de predicción} = \frac{\text{Predicciones correctas}}{n}, \quad \dots \text{Ecuación 23}$$



donde  $n$  es el número de observaciones muestrales.

Normalmente, el valor umbral que se suele adoptar para asignarle un valor a una predicción es 0.5; de tal manera que:

- Si  $\hat{Y}_i \geq 0.5$ , entonces se asigna a la predicción el valor 1;
- y si  $\hat{Y}_i < 0.5$ , se asigna a la predicción el valor 0.

A partir de aquí, se cuentan entonces los valores 1 y 0 asignados, comparándolos con los de las observaciones reales, para calcular de este modo el Pseudo  $R^2$  de predicción del modelo estimado (Ordaz et al., 2011).

#### • Comparación y elección entre modelos de elección discreta binaria

Los siguientes estadísticos que se va a tratar se refieren a “pérdida de información”. De algún modo sirven también para evaluar la bondad de un modelo, por cuanto se utilizan, en base a sus valores, para comparar las estimaciones realizadas por distintas modelizaciones y seleccionar de este modo la “mejor” (Ordaz et al., 2011).

- Akaike Information Criterion (AIC):

$$AIC = \frac{2k - 2 \ln(L)}{n} \quad \dots \text{Ecuación 24}$$

- Schwarz Criterion (SC):

$$SC = \frac{k \cdot \ln(n) - 2 \ln(L)}{n} \quad \dots \text{Ecuación 25}$$

- Hannan-Quinn Information Criterion (HQ):

$$HQ = \frac{2k \cdot \ln(\ln(n)) - 2 \ln(L)}{n} \quad \dots \text{Ecuación 26}$$

donde, para los tres estadísticos:

- k: número de regresores (incluido el término independiente);
- n: tamaño de la muestra;
- L: valor de la función de verosimilitud.

A la hora de establecer comparaciones entre distintos modelos, se considera con mejor ajuste aquél que presente unos valores más bajos en estos estadísticos. Además de estos estadísticos, también se podría emplear el estadístico de la razón de verosimilitud (LR-statistic), ya visto anteriormente (Ordaz et al., 2011).

## 2.2.12. Propensity score matching (PSM)

### 2.2.12.1. Marco de evaluación y conceptos básicos de emparejamiento

- **Modelo de Roy-Rubin**

La inferencia sobre el impacto de un tratamiento en el resultado de un individuo implica especular sobre cómo se habría desempeñado este individuo si no hubiera recibido el tratamiento (Roy, 1951; Rubin, 1974). El marco estándar en el análisis de evaluación para formalizar este problema es el enfoque de resultados potenciales o modelo de Roy-Rubin (Roy, 1951; Rubin, 1974).

Los pilares principales de este modelo son los individuos, el tratamiento y los resultados potenciales. En el caso de un tratamiento binario, el indicador de tratamiento  $D_i$  es igual a uno si el individuo  $i$  recibe tratamiento y cero en caso contrario y los resultados potenciales se definen entonces como  $Y_i(D_i)$  para cada individuo  $i$ , donde  $i = 1, \dots, N$  y  $N$  denota la población total, por lo tanto, el efecto del tratamiento para un individuo  $i$  se puede escribir como (Caliendo & Kopeinig, 2005):

$$\tau_i = Y_i(1) - Y_i(0). \quad \dots \text{Ecuación 27}$$

El problema fundamental de la evaluación surge porque sólo se observa uno de los resultados potenciales para cada individuo  $i$ . El resultado no observado se denomina resultado contrafáctico, en consecuencia, no es posible estimar el efecto del tratamiento individual  $\tau_i$ , por lo que se debe concentrarse en los efectos promedio del tratamiento poblacional (Caliendo & Kopeinig, 2005).

- **Parámetro de interés**

El parámetro que recibió mayor atención en la literatura de evaluación es el “Efecto Promedio del Tratamiento sobre los Tratados (ATT)”, que se define como (Caliendo & Kopeinig, 2005):

$$\tau_{ATT} = E(\tau|D = 1) = E[Y(1)|D = 1] - E[Y(0)|D = 1]. \quad \dots \text{Ecuación 28}$$

Como no se observa la media contrafactual para aquellos que reciben tratamiento  $E[Y(0)|D = 1]$ , es necesario elegir un sustituto adecuado para estimar el ATT. Usar el resultado medio de individuos no tratados  $E[Y(0)|D = 0]$  en estudios no experimentales generalmente no

es una buena idea, porque lo más probable es que los componentes que determinan la decisión de tratamiento también determinen la variable de resultado de interés. Por lo tanto, los resultados de los individuos del grupo de tratamiento y de comparación diferirían incluso en ausencia de tratamiento, lo que conduciría a un "sesgo de autoselección". Para el ATT se puede observar cómo (Caliendo & Kopeinig, 2005):

$$E[Y(1)|D = 1] - E[Y(0)|D = 0] = \tau_{ATT} + E[Y(0)|D = 1] - E[Y(0)|D = 0]. \dots \text{Ecuación 29}$$

La diferencia entre el lado izquierdo de la ecuación (29) y  $\tau_{ATT}$  es el llamado "sesgo de autoselección". El verdadero parámetro  $\tau_{ATT}$  sólo se identifica si (Caliendo & Kopeinig, 2005):

$$E[Y(0)|D = 1] - E[Y(0)|D = 0] = 0. \dots \text{Ecuación 30}$$

En los experimentos sociales en los que la asignación al tratamiento es aleatoria, esto se garantiza y se identifica el efecto del tratamiento. En los estudios no experimentales hay que invocar algunos supuestos de identificación para resolver el problema de sección planteado en la ecuación (29). Otro parámetro de interés es el "Efecto Promedio del Tratamiento" (ATE), que se define como (Caliendo & Kopeinig, 2005):

$$\tau_{ATE} = E[Y(1) - Y(0)]. \dots \text{Ecuación 31}$$

El desafío adicional al estimar el ATE es que se deben construir ambos resultados contrafactuales  $E[Y(1)|D = 0]$  y  $E[Y(0)|D = 1]$  (Caliendo & Kopeinig, 2005).

- **Supuesto de independencia condicional**



Una posible estrategia de identificación es suponer que, dado un conjunto de covariables  $X$  observables que no se ven afectadas por el tratamiento, los resultados potenciales son independientes de la asignación del tratamiento (Caliendo & Kopeinig, 2005):

$$(Sin\ confusión)\ Y(0), Y(1) \perp\!\!\!\perp D|X, \forall X. \quad \dots Ecuación\ 32$$

Esto implica que la selección se basa únicamente en características observables y que el investigador observa simultáneamente todas las variables que influyen en la asignación del tratamiento y los resultados potenciales (Caliendo & Kopeinig, 2005). Claramente, se trata de una suposición sólida y debe justificarse por la calidad de los datos disponibles. También debe quedar claro que el condicionamiento de todas las covariables relevantes está limitado en el caso de un vector  $X$  de alta dimensión. Por ejemplo, si  $X$  contiene  $s$  covariables que son todas dicotómicas, el número de coincidencias posibles será  $2^s$  (Caliendo & Kopeinig, 2005). Para abordar este problema de dimensionalidad, Rosenbaum & Rubin (1983) sugieren utilizar los llamados puntajes de equilibrio. Muestran que, si los resultados potenciales son independientes del tratamiento condicionado a las covariables  $X$ , también son independientes del tratamiento condicionado a una puntuación de equilibrio  $b(X)$ . La puntuación de propensión  $P(D = 1|X) = P(X)$ , es decir, la probabilidad de que un individuo participe en un tratamiento dadas sus covariables  $X$  observadas, es una posible puntuación de equilibrio. El supuesto de independencia condicional (CIA) basado en el

puntaje de propensión (PS) se puede escribir como (Caliendo & Kopeinig, 2005):

(Falta de confusión dada la PS)  $Y(0), Y(1) \perp\!\!\!\perp D | P(X), \forall X$ . ... Ecuación 33

- **Soporte común**

Otro requisito además de la independencia es el soporte común o condición de superposición. Descarta el fenómeno de perfecta previsibilidad de  $D$  dado  $X$ :

(Superposición)  $0 < P(D = 1 | X) < 1$ . ... Ecuación 34

Garantiza que las personas con los mismos valores de  $X$  tengan una probabilidad positiva de ser participantes y no participantes (Heckman et al., 1999)

- **Estrategia de estimación**

Dado que CIA cumple y suponiendo además que existe superposición entre ambos grupos, el estimador PSM para ATT se puede escribir en general como:

$\tau_{ATT}^{PSM} = E_{P(X)|D=1} \{E[Y(1)|D=1, P(X)] - E[Y(0)|D=0, P(X)]\}$ . ... Ecuación 35

Para decirlo en palabras, el estimador PSM es simplemente la diferencia media en los resultados sobre el apoyo común, ponderada adecuadamente por la distribución del puntaje de propensión de los participantes (Caliendo & Kopeinig, 2005).



### 2.2.12.2. Pasos para la implementación del emparejamiento por puntuación de propensión

- **Paso 1: Estimación del puntaje de propensión**

Al estimar el puntaje de propensión, se deben tomar dos decisiones. El primero se refiere al modelo a utilizar para la estimación y el segundo a las variables a incluir en este modelo (Caliendo & Kopeinig, 2005).

- Elección del modelo

Hay pocos consejos disponibles sobre qué forma funcional utilizar. En principio, se puede utilizar cualquier modelo de elección discreta. La preferencia por los modelos Logit o Probit (en comparación con los modelos de probabilidad lineal) se deriva de las conocidas deficiencias del modelo de probabilidad lineal, especialmente la improbabilidad de la forma funcional cuando la variable de respuesta está muy sesgada y las predicciones están fuera del [0,1] límites de probabilidades (Caliendo & Kopeinig, 2005). Sin embargo, cuando el propósito de un modelo es clasificar más que estimar coeficientes estructurales, resulta menos claro que estas críticas sean aplicables (Smith, 1997). Para el caso del tratamiento binario, donde se estima la probabilidad de participación versus no participación, los modelos Logit y Probit generalmente arrojan resultados similares. Por lo tanto, la elección no es demasiado crítica, aunque la distribución Logit tiene más densidad de masa en los límites (Caliendo & Kopeinig, 2005).

- Elección de variable

Hay más consejos disponibles sobre la inclusión (o exclusión) de covariables en el modelo de puntuación de propensión. La estrategia de emparejamiento se basa en el CIA y requiere que la(s) variable(s) de resultado sean independientes del tratamiento condicionado al puntaje de propensión. Por lo tanto, implementar el emparejamiento requiere elegir un conjunto de variables  $X$  que satisfagan de manera creíble esta condición (Caliendo & Kopeinig, 2005). Heckman et al. (1997) muestran que omitir variables importantes puede aumentar seriamente el sesgo en las estimaciones resultantes. Sólo se deben incluir variables que influyan simultáneamente en la decisión de participación y en la variable de resultado. Por lo tanto, la teoría económica, un conocimiento sólido de investigaciones previas y también información sobre los entornos institucionales deben guiar al investigador en la construcción del modelo (Sianesi, 2004).

- Método de acertar o fallar

El primero es el método de "acertar o fallar" o métrica de tasa de predicción, donde las variables se eligen para maximizar las tasas de predicción correcta dentro de la muestra. Este método clasifica una observación como "1" si el puntaje de propensión estimado es mayor que la proporción muestral de personas que reciben tratamiento, es decir,  $\hat{P}(X) > \bar{P}$ . En cambio, si  $\hat{P}(X) \leq \bar{P}$ , las observaciones se clasifican como "0". Este método maximiza la tasa de clasificación general de la muestra asumiendo que los costos de la clasificación errónea son iguales



para los dos grupos (Heckman et al., 1997). Pero claramente, debe tenerse en cuenta que el propósito principal de la estimación del puntaje de propensión no pretende predecir la selección hacia el tratamiento lo mejor posible sino equilibrar todas las covariables (Augurzky & Schmidt, 2000).

- Importancia estadística

El segundo enfoque se basa en la significancia estadística y es muy común en los libros de texto de econometría. Para ello se parte de una especificación parsimoniosa del modelo, por ejemplo, una constante, la edad y cierta información regional, y luego se "prueba" agregando variables de forma iterativa a la especificación. Se mantiene una nueva variable si es estadísticamente significativa a niveles convencionales. Si se combinan con el método de "acertar o fallar", las variables se mantienen si son estadísticamente significativas y aumentan las tasas de predicción en una cantidad sustancial (Heckman et al., 1998).

- Validación cruzada de dejar uno fuera

La validación cruzada de dejar uno fuera también se puede utilizar para elegir el conjunto de variables que se incluirán en la puntuación de propensión. Black & Smith (2003) implementan el procedimiento de selección de modelos comenzando con un modelo "mínimo" que contiene sólo dos variables. Posteriormente agregan bloques de variables adicionales y comparan los errores cuadráticos medios resultantes. Como nota de precaución, subrayan que esto equivale a elegir el modelo de puntuación de propensión basándose en consideraciones de bondad de ajuste, en lugar de basarse en la teoría y la evidencia sobre el conjunto de



variables relacionadas con la decisión de participación y los resultados (Black & Smith, 2003).

- **Paso 2: Elección de un algoritmo de coincidencia**

El estimador PSM en su forma general queda expresado en la ecuación (35). Todos los estimadores coincidentes contrastan el resultado de un individuo tratado con los resultados de los miembros del grupo de comparación. Los estimadores PSM difieren no sólo en la forma en que se define la vecindad para cada individuo tratado y se maneja el problema de apoyo común, sino también con respecto a las ponderaciones asignadas a estos vecinos. Existen diferentes algoritmos de PSM, algunos de ellos se muestran a continuación (Caliendo & Kopeinig, 2005):

- Coincidencia de vecinos más cercanos.
- Coincidencia de calibre y radio.
- Estratificación y coincidencia de intervalos.
- Coincidencia lineal local y de kernel.
- Coincidencia de vecino más cercano

El estimador de coincidencia más sencillo es la coincidencia de vecino más cercano (NN), donde el individuo del grupo de comparación se elige como compañero coincidente del individuo tratado más cercano en términos de puntuación de propensión, es decir, se proponen varias variantes de coincidencia de NN, por ejemplo, emparejamiento de NN “con reemplazo” y “sin reemplazo” (Caliendo & Kopeinig, 2005). En el primer caso, un individuo no tratado puede ser utilizado más de una vez



como compatible, mientras que en el segundo caso se considera sólo una vez, es decir, el emparejamiento con el reemplazo implica un equilibrio entre sesgo y varianza. Si se permite el reemplazo, la calidad promedio del emparejamiento aumentará y el sesgo disminuirá (Caliendo & Kopeinig, 2005). Esto es de particular interés con datos donde la distribución del puntaje de propensión es muy diferente en el grupo de tratamiento y de control (Caliendo & Kopeinig, 2005).

- Coincidencia de calibre y radio

La coincidencia de NN enfrenta el riesgo de malas coincidencias, si el vecino más cercano está lejos. Esto se puede evitar imponiendo un nivel de tolerancia a la distancia máxima del puntaje de propensión (calibrador). Imponer una pinza funciona en la misma dirección que permitir el reemplazo. Se evitan malas coincidencias y, por tanto, aumenta la calidad de las coincidencias. Sin embargo, si se pueden realizar menos coincidencias, la varianza de las estimaciones aumenta. La aplicación del calibrador significa que el individuo del grupo de comparación se elige como compañero de emparejamiento para un individuo tratado que se encuentra dentro del calibrador “rango de propensión” y es el más cercano en términos de puntuación de propensión (Caliendo & Kopeinig, 2005).

- Estratificación y emparejamiento de intervalos

La idea del emparejamiento de estratificación es dividir el soporte común del puntaje de propensión en un conjunto de intervalos (estratos) y calcular el impacto dentro de cada intervalo tomando la diferencia media en los resultados entre las observaciones tratadas y de control (Caliendo &



Kopeinig, 2005). Este método también se conoce como emparejamiento, bloqueo y subclasificación de intervalos (Rosenbaum & Rubin, 1983). Claramente, una pregunta que debe responderse es cuántos estratos deberían usarse en el análisis empírico. Cochrane & Chambers S. (1965) muestran que cinco subclases suelen ser suficientes para eliminar el 95% del sesgo asociado con una sola covariable.

- Emparejamiento lineal local y de núcleo

Los algoritmos de emparejamiento discutidos hasta ahora tienen en común que solo se utilizan unas pocas observaciones del grupo de comparación para construir el resultado contrafactual de un individuo tratado (Caliendo & Kopeinig, 2005). El emparejamiento kernel (KM) y el emparejamiento lineal local (LLM) son estimadores de emparejamiento no paramétricos que utilizan promedios ponderados de todos los individuos del grupo de control para construir el resultado contrafactual (Caliendo & Kopeinig, 2005). Por lo tanto, una ventaja importante de estos enfoques es la menor varianza que se logra porque se utiliza más información. Una desventaja de estos métodos es que posiblemente se utilicen observaciones que no coincidan (Caliendo & Kopeinig, 2005).

• **Paso 3: Verificación de la superposición o soporte común**

Se ha demostrado que ATT y ATE sólo se definen en la región de apoyo común. Por lo tanto, un paso importante es comprobar la superposición y la región de apoyo común entre el grupo de tratamiento y el de comparación. En la literatura se sugieren varias formas, siendo la más sencilla un análisis visual de la distribución de densidad de la puntuación



de propensión en ambos grupos. Sin embargo, existen dos métodos, donde el primero se basa esencialmente en comparar los mínimos y máximos del puntaje de propensión en ambos grupos y el segundo se basa en estimar la distribución de densidad en ambos grupos (Caliendo & Kopeinig, 2005). La implementación de la condición de apoyo común garantiza que cualquier combinación de características observadas en el grupo de tratamiento también pueda observarse en el grupo de control (Bryson et al., 2002).

- **Paso 4: Evaluación de la calidad de coincidencia**

Dado que no se condicionó a todas las covariables sino al puntaje de propensión, se debe verificar si el procedimiento de emparejamiento es capaz de equilibrar la distribución de las variables relevantes tanto en el grupo de control como en el de tratamiento. La idea básica de todos los enfoques es comparar la situación antes y después del emparejamiento y verificar si persisten diferencias después de condicionar la puntuación de propensión. Si hay diferencias, la coincidencia en la puntuación no fue (completamente) exitosa y se deben tomar medidas correctivas, por ejemplo, incluyendo términos de interacción en la estimación del puntaje de propensión (Caliendo & Kopeinig, 2005).

Un enfoque similar utiliza una prueba t de dos muestras para verificar si existen diferencias significativas en las medias de las covariables para ambos grupos (Rosenbaum & Rubin, 1983). Antes del emparejamiento se esperan diferencias, pero después del emparejamiento las covariables deben estar equilibradas en ambos grupos y, por lo tanto,

no deben encontrarse diferencias significativas (Caliendo & Kopeinig, 2005).

### 2.2.12.3. Supuestos del emparejamiento por puntuación de propensión

Al respecto, Heinrich et al. (2010) indica lo siguiente:

- **Supuesto 1 (Supuesto de Independencia Condicional o CIA)**

Existe un conjunto  $X$  de covariables, observables para el investigador, de modo que después de controlar estas covariables, los resultados potenciales son independientes del estado del tratamiento (Heinrich et al., 2010):

$$Y_1, Y_0 \perp D | X \quad \dots \text{Ecuación 36}$$

Esta es simplemente la notación matemática de la idea expresada en los párrafos anteriores, que establece: los resultados potenciales son independientes del estado del tratamiento, dado  $X$ . En otras palabras: después de controlar por  $X$ , la asignación del tratamiento es “casi aleatoria” (Heinrich et al., 2010).

- **Supuesto 2 (Condición de apoyo común)**

Para cada valor de  $X$ , existe una probabilidad positiva de ser tratado y no tratado (Heinrich et al., 2010):

$$0 < P(D = 1 | X) < 1 \quad \dots \text{Ecuación 37}$$



Esta última ecuación implica que la probabilidad de recibir tratamiento para cada valor de  $X$  se encuentra entre 0 y 1. Según las reglas de probabilidad, esto significa que la probabilidad de no recibir tratamiento se encuentra entre los mismos valores (Heinrich et al., 2010). Entonces, una forma sencilla de interpretar esta fórmula es la siguiente: la proporción de individuos tratados y no tratados debe ser mayor que cero para cada valor posible de  $X$  (Heinrich et al., 2010). El segundo requisito también se conoce como condición de superposición, porque asegura que haya suficiente superposición en las características de las unidades tratadas y no tratadas para encontrar coincidencias adecuadas (o un soporte común) (Heinrich et al., 2010).

## **2.3. MARCO CONCEPTUAL**

### **2.3.1. Empleo**

“Condición de las personas en edad y capacidad de trabajar que realizan algún tipo de trabajo, asalariado o no. Se refiere al grado de utilización de la fuerza laboral o de la población económicamente activa (PEA)” (Banco Central de Reserva del Perú [BCRP], 2011).

### **2.3.2. Empleo formal**

“Tradicionalmente, la Organización Internacional del Trabajo (OIT) define al empleo formal como el de las empresas de más de cinco trabajadores, así como a los profesionales y técnicos independientes” (BCRP, 2011). “De acuerdo a la Encuesta de Hogares del Ministerio de Trabajo, son empresas de



producción, comercio o servicios legalmente constituidas, que emplean trabajadores sin tener en cuenta el número de éstos” (BCRP, 2011).

### **2.3.3. Empleo informal**

“Empleo en empresas en las que los trabajadores no han firmado contrato laboral, no tienen seguro de trabajo y no están afiliados al sistema de pensiones” (BCRP, 2011). “Tradicionalmente se ha incluido al resto de independientes, a los que laboran en microempresas, a los trabajadores familiares no remunerados y a los empleados del hogar” (BCRP, 2011).

### **2.3.4. Desempleo**

“Condición de las personas en edad y disposición de trabajar que buscan activamente un puesto de trabajo, sin encontrarlo. También se denomina desempleo abierto.” (BCRP, 2011).

### **2.3.5. Población económicamente activa (PEA)**

Es la oferta de mano de obra en el mercado de trabajo y está constituida por el conjunto de personas, que contando con la edad mínima establecida (14 años en el caso del Perú), ofrecen la mano de obra disponible para la producción de bienes y/o servicios durante un período de referencia determinado (Instituto Nacional de Estadística e Informática [INEI], 2019). Por lo tanto, las personas son consideradas económicamente activas, si contribuyen o están disponibles para la producción de bienes y servicios. La PEA comprende a las personas, que durante el período de referencia estaban trabajando (ocupados) o buscando activamente un trabajo (desempleados) (INEI, 2019).



### **2.3.6. Población económicamente inactiva (PEI)**

“Es el grupo de personas en edad de trabajar que no participan en el mercado laboral. Es decir, que no realizan ni buscan alguna actividad económica” (INEI, 2019). “La PEI está conformada por los estudiantes, jubilados o pensionistas, rentistas, personas dedicadas a los quehaceres del hogar, ancianos, etc.” (INEI, 2019).

### **2.3.7 PEA ocupada**

Para determinar que una persona se encuentra ocupada se utilizan cuatro criterios: i) ocupados son las personas de 14 años y más de edad que estuvieron participando en alguna actividad económica, en el período de referencia; ii) los trabajadores dependientes, que teniendo empleo fijo, no trabajaron, la semana anterior, por hallarse de vacaciones, huelga, licencia por enfermedad, licencia pre y post-natal, etc., todas ellas pagadas; iii) los trabajadores independientes, que estuvieron temporalmente ausentes del trabajo durante el período de referencia; pero, la empresa o negocio siguió funcionando y iv) a las personas que no estuvieron en ninguna de las condiciones anteriores se les indaga si realizaron alguna actividad económica en el período de referencia, al menos una hora, por lo cual recibirá pago en dinero y/o especie (INEI, 2019).

### **2.3.8. PEA desocupada**

Según la OIT (1983), los desocupados son todas aquellas personas, de uno u otro sexo, que durante el período de referencia cumplen en forma simultánea con los tres requisitos siguientes: i) sin empleo, es decir, que no tienen ningún empleo, como asalariado o como independiente; ii) corrientemente disponible



para trabajar, es decir, con disponibilidad para trabajar en un empleo asalariado o independiente, durante el período de referencia y iii) en busca de empleo, es decir, que habían tomado acciones concretas para buscar un empleo asalariado o independiente, en un período de referencia especificado (INEI, 2019). “Este concepto abarca tanto a las personas que buscaron trabajo pero que trabajaron antes (cesantes), como a los que buscaron trabajo por primera vez (aspirantes)” (INEI, 2019).

### **2.3.9. Tasa de empleo**

“Es la proporción de personas que se encuentran ocupadas sobre el total de la población económicamente activa” (INEI, 2019).

### **2.3.10. Tasa de desempleo abierto**

“Es la proporción de personas desempleadas o desocupadas, que están buscando activamente un empleo, entre la PEA. Para determinar a los desempleados abiertos se consideran las condiciones siguientes: sin empleo, corrientemente disponible para trabajar y en busca de empleo” (INEI, 2019).

### **2.3.11. Ingresos**

“Recursos que se obtienen de modo regular como consecuencia de la actividad de una entidad” (BCRP, 2011).

### **2.3.12. Ingreso per cápita**

“Conjunto de remuneraciones promedio obtenida por los habitantes de un país en un periodo determinado, generalmente un año. Se utiliza para comparar estándares de vida entre países” (BCRP, 2011).



### **2.3.13. Sociedad de la información y el conocimiento (SIC)**

Son dos expresiones que en el campo educativo se refieren al uso de dispositivos digitales para facilitar el aprendizaje y consolidar un modelo integral de educación que cumpla con los objetivos tecno-pedagógicos de la actualidad (Pérez, 2018). En este sentido, la introducción de tecnologías en las aulas y el incremento de cursos en línea han abierto nuevos horizontes para mejorar la calidad de la educación y han incidido en la transformación de los modelos educativos que se basan en la infraestructura tecnológica y en el internet para procesar y transmitir información (Pérez, 2018).

### **2.3.14. Brecha digital**

En la SIC esta nueva desigualdad se ha denominado como “brecha digital”, la cual se reconoce como un fenómeno complejo que comprende aspectos políticos, económicos y sociales, y se relaciona íntimamente con problemas estructurales de la sociedad global como la pobreza, la exclusión, el desempleo, la precarización del trabajo, la inequidad en la distribución de la riqueza, además con otras problemáticas que surgen en la SIC marcada por la globalización (Reygadas, 2008, como se citó en Gómez et al., 2018).

### **2.3.15. Tecnologías de la información y la comunicación (TIC)**

Herramientas y procesos para acceder, recuperar, guardar, organizar, manipular, producir, intercambiar y presentar información por medios electrónicos; estos incluyen hardware, software y telecomunicaciones en la forma de computadores y programas tales como aplicaciones multimedia y sistemas de bases de datos (Sunkel, 2006, como se citó en Gómez et al., 2018).



### **2.3.16. Hardware**

“Es un sistema de computación actual, el cual consiste en una unidad de procesamiento central, almacenamiento primario, almacenamiento secundario, dispositivos de entrada, dispositivos de salida y dispositivos de comunicación” (Amaya, 2010).

### **2.3.17. Software**

Consiste en las instrucciones detalladas que controlan el funcionamiento de un sistema computacional, además, las funciones de software son: administrar los recursos computacionales de hardware, proporcionar las herramientas para aprovechar dichos recursos y actuar como intermediario entre las organizaciones y la información almacenada (Amaya, 2010).

### **2.3.18. Internet**

Es la implementación mejor conocida y la más grande de la interconexión de redes: el enlace de miles de redes individuales de todo el mundo, además, internet tiene una gama de capacidades que las organizaciones están usando para intercambiar información internamente o para comunicarse externamente con otras organizaciones, asimismo, esta gigantesca red de redes se ha convertido en un importante catalizador tanto para el comercio electrónico como para los negocios electrónicos (Amaya, 2010).

### **2.3.19. Beneficios de internet**

Internet, intranets y extranets se están convirtiendo en las plataformas principales para el comercio electrónico y los negocios electrónicos porque esta



tecnología ofrece muchos beneficios, además, la conectividad global de internet, su facilidad de uso, su bajo costo y sus capacidades multimedios, sirven para crear aplicaciones interactivas, servicios y productos, asimismo, con la tecnología internet las organizaciones pueden reducir los costos de comunicación y de transacción, mejorar la coordinación y la colaboración, y acelerar la distribución de conocimientos (Amaya, 2010).

## **2.4. HIPÓTESIS DE LA INVESTIGACIÓN**

### **2.4.1. Hipótesis general**

- Los factores que inciden en el acceso a internet son: el material de vivienda, el número de habitaciones, la energía eléctrica, el número de miembros, la existencia de niños, la existencia de estudiantes, la existencia de miembros con celular, la existencia de miembros con computadora y/o laptop, la región geográfica y la zona de ubicación. Mientras que los factores que inciden en el uso de internet son: la edad, el género, el estado civil, la educación, el tamaño de familia, la existencia de niños, la existencia de estudiantes, la disponibilidad de celular, la disponibilidad de computadora y/o laptop, la región geográfica y la zona de residencia; y consecuentemente el uso de internet genera un impacto positivo sobre el empleo en los hogares del Perú, en el año 2020.

### **2.4.2. Hipótesis específicas**

- Los factores que influyen en la probabilidad de que un hogar tenga acceso a internet disponible son: el material de vivienda, el número de habitaciones, la energía eléctrica, el número de miembros, la existencia de



niños, la existencia de estudiantes, la existencia de miembros con celular, la existencia de miembros con computadora y/o laptop, la región geográfica y la zona de ubicación.

- Los factores que influyen en la probabilidad de que un jefe de hogar use internet son: la edad, el género, el estado civil, la educación, el tamaño de familia, la existencia de niños, la existencia de estudiantes, la disponibilidad de celular, la disponibilidad de computadora y/o laptop, la región geográfica y la zona de residencia.
- El uso de internet genera un impacto positivo sobre el empleo.



## CAPÍTULO III

### MATERIALES Y MÉTODOS

#### 3.1. ENFOQUE DE INVESTIGACIÓN

Según Hernández et al. (2014), existen 3 enfoques de investigación: cuantitativo, cualitativo y mixto, los cuales constituyen posibles elecciones para enfrentar problemas de investigación y resultan igualmente valiosos. Sin embargo, en la presente investigación se hizo el uso del enfoque cuantitativo, ya que este enfoque “utiliza la recolección de datos para probar hipótesis con base en la medición numérica y el análisis estadístico, con el fin de establecer pautas de comportamiento y probar teorías” (Hernández et al., 2014).

#### 3.2. ALCANCE DE INVESTIGACIÓN

En la investigación cuantitativa existen 4 alcances de investigación: exploratorias, descriptivos, correlacionales y explicativos, los cuales son importantes para establecer los límites conceptuales y metodológicos de una investigación (Hernández et al., 2014). No obstante, en la presente investigación se utilizó el estudio de alcance explicativo, debido a que los “estudios explicativos pretenden establecer las causas de los sucesos o fenómenos que se estudian” (Hernández et al., 2014).

#### 3.3. DISEÑO DE INVESTIGACIÓN

El diseño de investigación es un plan o estrategia concebida para obtener la información que se desea con el fin de responder al planteamiento del problema. Existen 2 diseños de investigación: no experimentales y experimentales, en donde los estudios experimentales se clasifican en transversales y longitudinales, mientras que los estudios experimentales se clasifican en preexperimentos, cuasiexperimentos y experimentos



puros (Hernández et al., 2014). Sin embargo, para la presente investigación se utilizó el diseño de investigación experimental, específicamente el diseño cuasiexperimental. Los diseños cuasiexperimentales también manipulan deliberadamente, al menos, una variable independiente para observar su efecto sobre una o más variables dependientes, además, los sujetos no se asignan al azar a los grupos ni se emparejan, sino que dichos grupos ya están conformados antes del experimento (Hernández et al., 2014).

### **3.4. TIPO DE INVESTIGACIÓN**

Existen 2 tipos de investigación: básicas o aplicadas. Las investigaciones básicas son descriptivas, mientras que las investigaciones aplicadas son experimentales (R. Martínez & Espinal, 2023). No obstante, en el presente estudio se utilizó el tipo de investigación aplicada, debido a que se hizo el uso del diseño experimental, clasificado en el diseño cuasiexperimental, el cual implica que las variables serán modificadas para encontrar la solución al problema planteado.

### **3.5. MÉTODO DE INVESTIGACIÓN**

En una investigación existen 2 tipos de métodos: deductivo e inductivo. El método deductivo “se inicia con el análisis de los postulados, teoremas, leyes, principios, etc., de aplicación universal y de comprobada validez, para aplicarlos a soluciones o hechos particulares” (Rodríguez, 2007). Mientras que el método inductivo “se inicia con un estudio individual de los hechos y se formulan conclusiones universales que se postulan como leyes, principios o fundamentos de una teoría” (Rodríguez, 2007). En tal sentido, en el presente estudio se hizo el uso del método deductivo, debido a que está basado en la aplicación de una serie de teorías, leyes y principios.



### **3.6. UNIDAD DE ANÁLISIS**

La unidad de análisis se define como “quiénes o qué va a ser estudiado” (Vara, 2012). Es decir, “se trata de individuos, organizaciones, periodos, comunidades, situaciones, piezas producidas, eventos, etc.” (Hernández et al., 2014). Por lo tanto, para el presente estudio la unidad de análisis fueron los hogares del Perú que existieron en el año 2020.

### **3.7. POBLACIÓN Y MUESTRA**

#### **3.7.1. Población**

La población “es el conjunto de todos los individuos (objetos, personas, documentos, data, eventos, empresas, situaciones, etc.) a investigar” (Vara, 2012). En ese contexto, la población total para la presente investigación fue constituida por los hogares del Perú que existieron en el año 2020, el cual según el Instituto Nacional de Estadística e Informática (2020) consta de 14,742,326 hogares aproximadamente.

#### **3.7.2. Muestra**

“La muestra es un subgrupo de la población de interés sobre el cual se recolectarán datos, y que tiene que definirse y delimitarse de antemano con precisión, además de que debe ser representativo de la población” (Hernández et al., 2014). Por consiguiente, la muestra para el presente estudio fue el subgrupo de la población conformado por los hogares del Perú en el año 2020, el cual según el INEI (2020) consta de 53,423 hogares. Sin embargo, esta muestra fue reducida a 26,663 hogares para los objetivos específicos 1 y 2, ya que existió datos perdidos por el simple hecho de que muchos hogares respondieron de manera incompleta

los cuestionarios, algunos simplemente se abstuvieron a responder la encuesta, otros estuvieron ausentes y por otras razones más, mientras que para el objetivo específico 3 se redujo a 26,483 hogares, ya que para estimar el impacto del uso de internet sobre el empleo se trabajó exclusivamente con la PEA.

### 3.8. TIPO DE MUESTREO

En la investigación cuantitativa existen 2 tipos de muestro: probabilístico y no probabilístico. El primero conocido también como “muestreo aleatorio”, utiliza el azar y las estadísticas para determinar el tamaño y la selección de cada integrante de la muestra, por consiguiente, cumple con el principio de equi-probabilidad, ya que los elementos de la población tienen la misma probabilidad de ser elegidos en una muestra, en cambio, el segundo no se basa en el principio de la equi-probabilidad, ya que son técnicas que siguen otros criterios de selección, procurando que la muestra obtenida sea lo más representativa posible (Vara, 2012). En ese contexto, según el INEI (2020) el muestreo con la que trabajó fue del tipo probabilística, de áreas, estratificada, multietápica e independiente en cada departamento de estudio. Para determinar el tamaño de la muestra, el INEI ha utilizado la siguiente fórmula de población finita:

$$n = \frac{Z^2 \times p \times q \times N}{E^2 \times (N - 1) + Z^2 \times p \times q} \quad \dots \text{Ecuación 38}$$

Donde:

- n: Es el tamaño de la muestra estimada.
- N: Es el tamaño de la población objetivo.
- p: Es la proporción de la población ocupada.
- q: Es la proporción de la población no ocupada.
- E: Es el error permisible al 5%.



- Z: Es el valor de la desviación estándar, a un nivel de confianza del 95%.

### **3.9. TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS**

#### **3.9.1. Técnicas de recolección de datos**

La técnica de recolección de datos que ha utilizado el INEI es la encuesta, sin embargo, para el presente estudio se hizo el uso de la técnica de análisis de datos secundarios de registros públicos o análisis documental, debido a que se realizó una revisión y selección exhaustivo de los datos proporcionados por el INEI, a través de la ENAHO.

#### **3.9.2. Instrumentos de recolección de datos**

El instrumento es el “recurso que utiliza el investigador para registrar información o datos sobre las variables que tiene en mente” (Hernández et al., 2014). Por consiguiente, el instrumento de recolección de datos que ha utilizado el INEI es el cuestionario estructurado con 401 preguntas y 329 ítems, sin embargo, para el presente estudio el instrumento de recolección de datos fue la ficha bibliográfica.

### **3.10. FUENTES DE INFORMACIÓN Y PROCEDIMIENTO DE ANÁLISIS DE DATOS**

#### **3.10.1. Fuentes de información**

La información que se utilizó para esta investigación es una información secundaria, debido a que se extrajo la base de datos de la ENAHO proporcionado por el INEI. Cabe aclarar que, “la ENAHO es la investigación que permite al INEI



desde el año 1995, efectuar el seguimiento de los indicadores sobre las condiciones de vida” (INEI, 2020).

### **3.10.2. Procedimiento de análisis de datos**

El procedimiento para realizar el análisis de datos comenzó con la recopilación de la información disponible en la plataforma digital del INEI en base a la ENAHO sobre las condiciones de vida y pobreza, y terminó con la contrastación de la hipótesis de esta investigación. Los pasos que se siguieron de manera detallada para realizar el análisis de datos se muestran a continuación:

- Búsqueda de información en la plataforma digital del INEI.
- Selección de preguntas y módulos referente al presente estudio.
- Creación de variables e indicadores.
- Estimación del modelo Logit y Probit.
- Elección del mejor modelo econométrico.
- Estimación de efectos marginales.
- Estimación de impacto sobre la variable de resultado.
- Presentación de resultados.
- Contrastación de las hipótesis.

Para cumplir con el procedimiento mencionado anteriormente se ha utilizado el software estadístico Stata/MP versión 16.0 que permite el manejo y la gestión de la base de datos, el análisis estadístico, el trazado de gráficos y las simulaciones de diferentes modelos econométricos de un contexto real. Complementariamente, se ha utilizado el software Microsoft Excel 2021 para presentar, organizar, graficar, recopilar y describir los resultados, además, se ha

utilizado el software Microsoft Word 2021 para sintetizar y redactar los resultados obtenidos en la investigación.

### 3.11. IDENTIFICACIÓN Y DESCRIPCIÓN DE VARIABLES

Para la construcción y selección de variables se acudió a la base de datos del INEI que se encuentra en su plataforma digital, en la sección de microdatos. Una vez ingresado a dicha plataforma se seleccionó la Encuesta Nacional de Hogares con la metodología actualizada y se descargó información referente a las características de la vivienda y del hogar, características de los miembros del hogar, educación, empleo e ingresos, equipamiento del hogar y sumaria (variables calculadas).

Después de haber extraído la base de datos del INEI se procedió a seleccionar las variables preliminares que sirvieron para generar y crear las variables definitivas. (ver Tabla N° 1).

**Tabla 1**

*Descripción de las variables preliminares*

<b>Código del módulo</b>	<b>Módulo</b>	<b>Variable</b>	<b>Descripción</b>
100	Características de la vivienda y del hogar	p102	El material predominante en las paredes exteriores es.
		p103	El material predominante en los pisos es.
		p103a	El material predominante en los techos es.
		p104	¿Cuántas habitaciones en total tiene la vivienda, sin contar el baño, la cocina, los pasadizos, ni el garaje?
		p1121	Tipo de alumbrado del hogar: electricidad.
		p1142	Su hogar tiene: ¿teléfono celular?
		p1144	Su hogar tiene: ¿Conexión a internet?
200	Características de los miembros del hogar	p207	p207: Sexo.
		p208a	¿Qué edad tiene en años cumplidos?
		p209	¿Cuál es su estado civil o conyugal?
		dominio	Dominio geográfico
		estrato	Estrato geográfico

Código del módulo	Módulo	Variable	Descripción
300	Educación	p303	El año pasado (...), ¿estuvo matriculado en algún centro o programa de educación básica o superior?
		p306	Este año, ¿está matriculado en algún centro o programa de educación básica o superior?
		p301a	¿Cuál es el último año o grado de estudios y nivel que aprobó? – nivel.
		p314a	En el mes anterior, ¿Ud. hizo uso del servicio de internet?
500	Empleo e ingresos	ocu500	Indicador de la pea.
612	Equipamiento del hogar	p612n	Equipamiento del hogar.
		p612	¿Su hogar tiene?
34	Sumarias (variables calculadas)	mieperho	Total de miembros del hogar.
		inghog2d	Ingreso neto total.

Fuente: Elaboración propia, en base a la información del ENAHO - INEI

Luego de seleccionar las variables preliminares de los distintos módulos se procedió a generar y adecuar las variables para cada objetivo específico de la investigación, a través del software estadístico StataMP 16.0.

### 3.11.1. Descripción de las variables para el objetivo específico 1

En la Tabla N° 2 se muestra las variables que fueron utilizado para alcanzar el específico objetivo 1. Todas estas variables fueron generadas mediante el software estadístico StataMP 16.0, en base a las variables preliminares que se encuentran en los módulos 100, 200, 300, 612 y 34. En ese contexto, las variables material, estudiantes y computadora fueron generado a partir de una combinación de 2 a 3 variables preliminares, estos son: p102, p103y p103a; p303 y p306; p612n y p612, respectivamente. Sin embargo, las demás variables fueron generadas a partir de una sola variable preliminar cada una, estos son: p104, p1121, p208a, p1142, p1144, dominio y estrato. Cabe mencionar que las variables costa y rural

no aparecen en el modelo econométrico, ni mucho menos en la matriz de operacionalización de variables, debido a que se consideraron como las variables de base, es decir, son variables con las que se hicieron la comparación al momento de interpretar los impactos marginales.

**Tabla 2**

*Descripción de las variables definitivas para el O.E.1.*

<b>Tipo de variable</b>	<b>Nombre de la variable</b>	<b>Descripción de la variable</b>	
Variable dependiente	Acceso a internet	Acceso a conexión de internet disponible en el hogar.	
	Material	Material con la que fue construida el hogar.	
	Habitaciones	Cantidad de habitaciones que hay en el hogar.	
	Electricidad	Disponibilidad de la energía eléctrica en el hogar.	
	Miembros	Cantidad de miembros que hay en el hogar.	
	Niños	Existencia de niños menores a 6 años en el hogar.	
	Estudiantes	Existencia de estudiantes en el hogar.	
	Variable independiente	Celular	Disponibilidad de celular por algún miembro del hogar.
		Computadora	Disponibilidad de computadora por algún miembro del hogar.
		Costa	Ubicación geográfica del hogar - costa.
Sierra		Ubicación geográfica del hogar - sierra.	
Selva		Ubicación geográfica del hogar - selva.	
Urbano		Zona de residencia del hogar - urbano.	
Rural	Zona de residencia del hogar - rural.		

Fuente: Elaboración propia, en base a la información del ENAHO - INEI

### **3.11.2. Descripción de las variables para el objetivo específico 2**

En la Tabla N° 3 se muestra las variables que fueron utilizado para alcanzar el objetivo específico 2. Todas estas variables fueron generadas mediante el software estadístico StataMP 16.0, en base a las variables preliminares que existen en los módulos 200, 300, 612 y 34. En ese sentido, las variables estudiantes y computadora fueron generado a partir de una combinación de 2 variables preliminares, estos son: p303 y p306; p612n y p612, respectivamente. En cambio,

las demás variables fueron generadas en base a una sola variable preliminar cada una, estos son: p208a, p207, p209, p301a, p208a, p1142, p314a, mieperho, dominio y estrato. Cabe aclarar que las variables analfabeto, costa y rural, no aparecen en el modelo econométrico, ni mucho menos en la matriz de operacionalización de variables, debido a que se consideraron como las variables de base, es decir, son variables con las que se hicieron la comparación al momento de interpretar los impactos marginales.

**Tabla 3**

*Descripción de las variables definitivas para el O.E.2.*

<b>Tipo de variable</b>	<b>Nombre de la variable</b>	<b>Descripción de la variable</b>
Variable dependiente	Uso de internet	Uso de internet por parte del jefe de hogar.
	Edad	Edad del jefe de hogar.
	Genero	Sexo del jefe de hogar.
	Estado civil	Estado civil del jefe de hogar.
	Analfabeto	Educación del jefe de hogar sin nivel.
	Primaria	Educación del jefe de hogar con nivel de primaria completa o incompleta.
	Secundaria	Educación del jefe de hogar con nivel de secundaria completa o incompleta.
	Superior no universitaria	Educación del jefe de hogar con nivel de superior no universitario completa o incompleta.
	Superior universitaria	Educación del jefe de hogar con nivel de superior universitario completa o incompleta.
	Maestría / Doctorado	Educación del jefe de hogar con nivel de maestría o doctorado.
Variable independiente	Miembros	Cantidad de miembros que hay en el hogar.
	Niños	Existencia de niños menores a 6 años en el hogar.
	Estudiantes	Existencia de estudiantes en el hogar.
	Celular	Disponibilidad de celular por parte del jefe de hogar.
	Computadora	Disponibilidad de computadora por parte del jefe de hogar.
	Costa	Región geográfica donde vive el jefe de hogar – costa.
	Sierra	Región geográfica donde vive el jefe de hogar – sierra.
	Selva	Región geográfica donde vive el jefe de hogar – selva.

<b>Tipo de variable</b>	<b>Nombre de la variable</b>	<b>Descripción de la variable</b>
	Urbano	Zona donde reside el jefe de hogar – urbano.
	Rural	Zona donde reside el jefe de hogar – rural.

Fuente: Elaboración propia, en base a la información del ENAHO - INEI

### 3.11.3. Descripción de las variables para el objetivo específico 3

Dado que el objetivo específico 3 de la presente investigación está relacionada con la evaluación de impacto, entonces se hace mención a 3 tipos de variables: control, tratamiento y resultado. En tal sentido, la Tabla N° 4 muestra las variables que fueron utilizado para alcanzar este objetivo, las cuales fueron generadas en base a las variables preliminares que se encuentran en los módulos 200, 300, 612 y 34, a través del software estadístico StataMP 16.0. Las variables ingreso per cápita, estudiantes y computadora fueron generado a partir de una combinación de 2 variables preliminares, las cuales son: inghog2d y mieperho; p303 y p306; p612n y p612, respectivamente, en cambio, las demás variables fueron generadas en base a una sola variable preliminar cada una, las cuales son: p208a, p207, p209, p301a, p208a, p1142, p314a, mieperho, dominio y estrato. Cabe aclarar que, la variable ingreso per cápita fue utilizada como un indicador o una variable proxi de la variable empleo.

**Tabla 4**

*Descripción de las variables definitivas para el O.E.3.*

<b>Tipo de variable</b>	<b>Nombre de la variable</b>	<b>Descripción de la variable</b>
Variable de resultado	Ingreso per cápita	Ingreso neto per cápita del jefe de hogar, en términos de meses.
Variable de tratamiento	Uso de internet	Uso de internet por parte del jefe de hogar.
Variables de control	Edad	Edad del jefe de hogar.
	Genero	Sexo del jefe de hogar.
	Estado civil	Estado civil del jefe de hogar.



<b>Tipo de variable</b>	<b>Nombre de la variable</b>	<b>Descripción de la variable</b>
	Analfabeto	Educación del jefe de hogar sin nivel.
	Primaria	Educación del jefe de hogar con nivel de primaria completa o incompleta.
	Secundaria	Educación del jefe de hogar con nivel de secundaria completa o incompleta.
	Superior no universitaria	Educación del jefe de hogar con nivel de superior no universitario completa o incompleta.
	Superior universitaria	Educación del jefe de hogar con nivel de superior universitario completa o incompleta.
	Maestría / Doctorado	Educación del jefe de hogar con nivel de maestría o doctorado.
	Miembros	Cantidad de miembros que hay en el hogar.
	Niños	Existencia de niños menores a 6 años en el hogar.
	Estudiantes	Existencia de estudiantes en el hogar.
	Celular	Disponibilidad de celular por parte del jefe de hogar.
	Computadora	Disponibilidad de computadora por parte del jefe de hogar.
	Costa	Región geográfica donde vive el jefe de hogar – costa.
	Sierra	Región geográfica donde vive el jefe de hogar – sierra.
	Selva	Región geográfica donde vive el jefe de hogar – selva.
	Urbano	Zona donde reside el jefe de hogar – urbano.
	Rural	Zona donde reside el jefe de hogar – rural.

Fuente: Elaboración propia, en base a la información del ENAHO - INEI

### **3.12. ESPECIFICACIÓN DEL MODELO ECONOMÉTRICO**

#### **3.12.1. Especificación del modelo econométrico para el objetivo específico 1**

Para especificar el modelo econométrico sobre el acceso a internet se utilizó el enfoque de Grazzi (2011), que consiste en estimar la probabilidad de que un individuo viva en un hogar con un ordenador conectado a internet. El punto interesante es que la decisión de tener o no una conexión a internet disponible en el hogar es una elección que se refiere al hogar en su conjunto, mientras que la decisión de utilizar internet es individual. En ese contexto, el modelo econométrico tiene variables referidas al hogar en el que vive una persona y, por

lo tanto, comunes a todos los miembros del hogar, en lugar de ser idiosincrásicas individualmente, tal como se muestra a continuación:

$$\begin{aligned} Pr(\text{Acceso a internet} = 1) &= F[\alpha_0 + \alpha_1(\text{Material}) + \alpha_2(\text{Habitaciones}) \\ &+ \alpha_3(\text{Electricidad}) + \alpha_4(\text{Miembros}) + \alpha_5(\text{Niños}) \\ &+ \alpha_6(\text{Estudiantes}) + \alpha_7(\text{Celular}) + \alpha_8(\text{Computadora}) \\ &+ \alpha_9(\text{Sierra}) + \alpha_{10}(\text{Selva}) + \alpha_{11}(\text{Urbano}) + \mu_i] \end{aligned}$$

... Ecuación 39

Donde:

- ***Pr (Acceso a internet = 1)***: Es la probabilidad de que un hogar tenga conexión a internet disponible.
- ***Material***: Es una covariable que representa el material con la que fue construida el hogar y se clasifica en 2 categorías, rustico y no rustico.
- ***Habitaciones***: Es una covariable que representa el número de habitaciones con la que cuenta el hogar.
- ***Electricidad***: Es una covariable que representa si el hogar cuenta con energía eléctrica o no.
- ***Miembros***: Es una covariable que representa el número de miembros que hay en el hogar.
- ***Niños***: Es una covariable que representa si en el hogar existen niños menores a 6 años o no.
- ***Estudiantes***: Es una covariable que representa si en el hogar existen estudiantes o no.



- **Celular:** Es una covariable que representa si en el hogar hay algún miembro que tiene teléfono móvil o no.
- **Computadora:** Es una covariable que representa si en el hogar hay algún miembro que tiene computadora y/o laptop o no.
- **Sierra:** Es una covariable que representa si el hogar se encuentra ubicada en la sierra o en la costa.
- **Selva:** Es una covariable que representa si el hogar se encuentra ubicada en la selva o en la costa.
- **Urbano:** Es una covariable que representa si el hogar se encuentra ubicada en zona urbana o rural.
- $\mu_i$ : Es el error estocástico o las covariables no observables.

### 3.12.2. Especificación del modelo econométrico para el objetivo específico 2

Para especificar el modelo econométrico sobre el uso del internet se utilizó el enfoque de Grazzi (2011), donde muestra un modelo empírico para analizar los determinantes y patrones de uso de internet en los países de América Latina. Para tal fin, supone que la decisión de un individuo de usar o no internet sigue un marco estándar de maximización de utilidad, donde un individuo elige usar internet cuando obtiene una utilidad positiva, es decir, cuando los beneficios asociados con su uso exceden sus costos.

En ese contexto, la utilidad derivada del uso de internet ( $U_{i,U}$ ) o no uso ( $U_{i,N}$ ) se modela como una función lineal de un vector de características socioeconómicas del individuo ( $X_i$ ) y un término estocástico aditivo que representa los no observables y el error de medición ( $\varepsilon_i$ ):

$$U_{i,U} = \alpha_U + X_i\beta_U + \varepsilon_{i,U} \quad \dots \text{Ecuación 40}$$

$$U_{i,N} = \alpha_N + X_i\beta_N + \varepsilon_{i,N} \quad \dots \text{Ecuación 41}$$

En consecuencia, la probabilidad de que el *i*-ésimo individuo utilice internet es la probabilidad de que su utilidad de uso sea mayor que su utilidad de no uso:

$$\Pr(\text{Internet Use} = 1) = \Pr(U_{i,U} > U_{i,N}) = F[(\alpha_U - \alpha_N) + X_i(\beta_U - \beta_N) \dots \text{Ecuación 42}$$

Donde *F* es la función de distribución acumulada del término de error ( $\varepsilon_{i,U} - \varepsilon_{i,N}$ ). Si se asume que ( $\varepsilon_{i,U} - \varepsilon_{i,N}$ ) se distribuye normalmente, el modelo se puede estimar mediante la estimación de máxima verosimilitud (MLE) con una regresión Logit o Probit simple. De manera empírica, se estima la siguiente ecuación para modelar la probabilidad de que un individuo use internet:

$$\begin{aligned} \Pr(\text{Uso de internet} = 1) &= F[\beta_0 + \beta_1(\text{Edad}) + \beta_2(\text{Genero}) + \beta_3(\text{Est_civil}) \\ &+ \beta_4(\text{Primaria}) + \beta_5(\text{Secundaria}) + \beta_6(\text{Sup_no_univ}) \\ &+ \beta_7(\text{Sup_univ}) + \beta_8(\text{Maestria}) + \beta_9(\text{Miembros}) \\ &+ \beta_{10}(\text{Niños}) + \beta_{11}(\text{Estudiantes}) + \beta_{12}(\text{Celular}) \\ &+ \beta_{13}(\text{Computadora}) + \beta_{14}(\text{Sierra}) + \beta_{15}(\text{Selva}) \\ &+ \beta_{16}(\text{Urbano}) + \varepsilon_i] \end{aligned}$$

... Ecuación 43

Donde:

- ***Pr(Uso de internet = 1)***: Es la probabilidad de que el jefe de hogar use internet.



- **Edad:** Es una covariable que representa la edad del jefe de hogar en años cumplidos.
- **Genero:** Es una covariable que representa el sexo del jefe de hogar y se clasifica en 2 categorías, hombre y mujer.
- **Est\_civil:** Es una covariable que representa el estado civil del jefe de hogar y se clasifica en 2 categorías, casado y no casado.
- **Primaria:** Es una covariable que representa si el jefe de hogar ha logrado estudiar la educación primaria o no.
- **Secundaria:** Es una covariable que representa si el jefe de hogar ha logrado estudiar la educación secundaria o no.
- **Sup\_no\_univ:** Es una covariable que representa si el jefe de hogar ha logrado estudiar la educación superior no universitaria o no.
- **Sup\_univ:** Es una covariable que representa si el jefe de hogar ha logrado estudiar la educación superior universitaria o no.
- **Maestria:** Es una covariable que representa si el jefe de hogar ha logrado estudiar una maestría y doctorado o no.
- **Miembros:** Es una covariable que representa el tamaño de familia con la que cuenta el jefe de hogar.
- **Niños:** Es una covariable que representa si el jefe de hogar tiene niños menores a 6 años o no.
- **Estudiantes:** Es una covariable que representa si el jefe hogar es estudiante y/o tiene algún miembro que está estudiando o no.
- **Celular:** Es una covariable que representa si el jefe de hogar cuenta con celular o no.



- **Computadora:** Es una covariable que representa si el jefe de hogar cuenta con computadora y/o laptop o no.
- **Sierra:** Es una covariable que representa si el jefe de hogar vive en la sierra o en la costa.
- **Selva:** Es una covariable que representa si el jefe de hogar vive en la selva o en la costa.
- **Urbano:** Es una covariable que representa si el jefe de hogar reside en la zona urbana o rural.
- $\varepsilon_i$ : Es el error estocástico o las covariables no observables.

### 3.12.3. Especificación del modelo econométrico para el objetivo específico 3

En este caso, para estimar el impacto del uso de internet sobre el empleo se optó por usar el enfoque empírico desarrollado por Navarro (2010), donde plantea cuál habría sido el desempeño de los individuos si no usaran internet. Dado que la decisión de adoptar internet no es aleatoria, no es posible observar el resultado para los individuos que no usan internet porque eso incurriría en un sesgo de selección. Sin embargo, se debe crear un contrafactual adecuado del resultado de los usuarios condicionados a la no utilización. Se pueden utilizar diferentes técnicas para abordar este problema. En este caso, se implementó el método de Propensity Score Matching (PSM).

En este contexto, el tratamiento entonces es una variable ficticia  $U_i$  (usuarios) que toma un valor de 1 si el individuo usa internet y cero en caso contrario. Los valores de  $U_i$  determinan la asignación de los individuos a los grupos de tratamiento y control, respectivamente. Sea  $Y_{1i}$  el resultado del individuo  $i$  como resultado del tratamiento. El efecto causal del uso de internet

sobre el resultado del individuo tratado entonces es  $Y_{1i} - Y_{0i}$ , donde  $Y_{0i}$  es el resultado evaluado en caso de no uso ( $U_i = 0$ ). Claramente,  $Y_{0i}$  no es observable. Es estándar definir el efecto promedio del tratamiento sobre la variable de resultado como:

$$E(Y_{1i} - Y_{0i}|U_i = 1) = E(Y_{1i}|U_i = 1) - E(Y_{0i}|U_i = 1) \quad \dots \text{Ecuación 44}$$

Si bien se observa el primer término, no se observa el segundo. Un estimador de este contrafactual ampliamente utilizado en la literatura de evaluación es,

$$E(Y_{0i}|U_i = 1) = E(Y_{0i}|P(X), U_i = 1) = E(Y_{0i}|P(X), U_i = 0), \quad \dots \text{Ecuación 45}$$

donde  $P(X)$  es la probabilidad de uso de internet condicionada a un conjunto de características observables  $X$ . Nótese que el valor promedio del resultado debe ser independiente del indicador de tratamiento (independencia condicional). También se debe considerar un rango para  $P(X)$ , tal que la comparación de los valores esperados entre los grupos de control y tratamiento sea factible (apoyo común).

En consecuencia, se estima primero un modelo Logit o Probit para la probabilidad de uso de internet (puntaje de propensión) condicionado a un conjunto de variables observables  $X$ . Luego se necesita encontrar un grupo de control muy similar al grupo de tratamiento en términos de su probabilidad predicha de uso de internet ( $p_i$ ). Esto requiere elegir un conjunto  $X$  de variables que no estén influenciadas por el tratamiento (Todd, 1999), es decir, características previas al tratamiento. Para lo cual, los elementos de  $X$  deben

incluir variables que se cree que afectan la probabilidad de uso, pero no el resultado.

Según Todd (2008), no existe una base teórica sobre cómo elegir  $X$  y qué variables se incluyen en  $X$ , las cuales pueden tener implicaciones importantes para el desempeño del estimador. Rosenbaum & Rubin (1983) proponen como prueba de especificación (equilibrio) elegir un conjunto  $X$  tal que no haya diferencias en  $X$  entre los dos grupos después del condicionamiento para  $P(X)$ . Sin embargo, en este estudio se siguió el procedimiento del `psmatch2` de Stata. Una vez que se estimaron los puntajes de propensión, se emparejó los grupos usando los métodos del vecino más cercano, radio y calibre y kernel. Es decir, para cada usuario con puntuación de propensión  $p_i$ , se seleccionó un individuo  $j$  tal que su puntuación de propensión  $p_j$  sea lo más cercana posible a  $p_i$ . Después de este procedimiento, se emparejó grupos de usuarios y no usuarios. Finalmente, se calculó el efecto del uso de internet comparando los resultados de los dos grupos de observaciones emparejadas. Como se conoce comúnmente en la literatura de evaluación, este es el Efecto Promedio de Tratamiento en los Tratados ( $ATT$ ), el cual para el presente estudio se representa de la siguiente manera:

$$ATT = E(Y_{1i}|P(X), U_i = 1) - E(Y_{0i}|P(X), U_i = 0) \quad \dots Ecuación 46$$

$$ATT = F[Y_i, P(X)] \quad \dots Ecuación 47$$

Donde:

- **$ATT$** : Es el efecto promedio de tratamiento en los tratados (valor del impacto del uso de internet sobre el ingreso per cápita)



- $E(Y_{1i}|P(X), U_i = 1)$ : Es el valor esperado del resultado en presencia del tratamiento para el grupo de tratados (resultado potencial en relación al ingreso per cápita de los jefes de hogar que usan internet).
- $E(Y_{0i}|P(X), U_i = 0)$ : Es el valor esperado del resultado en ausencia del tratamiento para el grupo de no tratados (resultado potencial en relación al ingreso per cápita de los jefes de hogar que no usan internet).
- $Y_i$ : Es una variable de resultado que representa el ingreso per cápita del jefe del hogar, en términos de meses.
- $P(X)$ : Es una variable de tratamiento que representa la probabilidad de que el jefe de hogar use internet, condicionada a un conjunto de características observables X (variables de control).

### 3.13. OPERACIONALIZACIÓN DE VARIABLES

Para los objetivos específicos 1 y 2, las variables se clasificaron en variables dependientes e independientes, en cambio, para el objetivo específico 3, las variables se clasificaron en variables de resultado, tratamiento y control, de acuerdo a la naturaleza de su análisis.

#### 3.13.1. Operacionalización de variables para el objetivo específico 1

Para este objetivo específico las variables se clasificaron de la siguiente manera (ver Tabla N° 5):

- Variable dependiente: Acceso a internet.
- Variable independiente: Factores de acceso a internet.

**Tabla 5**

*Resumen de las variables para el O.E.1.*

Variables	Definición conceptual	Definición operacional	Dimensiones	Indicadores	Ítems
<b>Variable 1:</b> <b>Acceso a internet</b> (Variable dependiente)	Es la disposición o decisión de los miembros de un hogar para hacerse conectar el internet o no desde una red pública hacia su hogar.	Esta variable fue medida con el cuestionario de la Encuesta Nacional de Hogares (ENAHG), a través del módulo 100.	Acceso a internet	1=Tiene conexión a internet 0=No tiene conexión a internet	<b>p1144:</b> Su hogar tiene: ¿Conexión a internet?
			Material de vivienda	1=Rústico 0=No rústico	<b>p102:</b> El material predominante en las paredes exteriores es. <b>p103:</b> El material predominante en los pisos es. <b>p103a:</b> El material predominante en los techos es.
<b>Variable 2:</b> <b>Factores de acceso a internet</b> (Variable independiente)	Los factores de acceso a internet son las covariables que explican el simple hecho de que los miembros de un hogar decidan o no hacerse instalar la conexión de internet disponible en su hogar, siempre en cuando la red o infraestructura de internet exista en el lugar donde este ubicada la vivienda.	Esta variable fue medida con el cuestionario de la Encuesta Nacional de Hogares (ENAHG), a través de los módulos 100, 200, 300, 612 y 34.	Habitaciones	Número de habitaciones en el hogar	<b>p104:</b> ¿Cuántas habitaciones en total tiene la vivienda, sin contar el baño, la cocina, los pasadizos, ni el garaje?
			Energía eléctrica	1=Si 0=No	<b>p1121:</b> Tipo de alumbrado del hogar: electricidad.
			Miembros	Número de miembros en el hogar	<b>mieperho:</b> Total de miembros del hogar.
			Niños	1=Si 0=No	<b>p208a:</b> ¿Qué edad tiene en años cumplidos?
			Estudiantes	1=Si 0=No	<b>p303:</b> El año pasado (...), ¿estuvo matriculado en algún centro o programa de educación básica o superior? <b>p306:</b> Este año, ¿está matriculado en algún centro o programa de educación básica o superior?
			Celular	1=Si 0=No	<b>p1142:</b> Su hogar tiene: ¿teléfono celular?
			Computadora	1=Si 0=No	<b>p612n:</b> Equipamiento del hogar. <b>p612:</b> ¿Su hogar tiene?
			Región geográfica	1=Costa 2=Sierra 3=Selva	<b>dominio:</b> Dominio geográfico.
Zona de residencia	1=Zona Urbana 0=Zona rural	<b>estrato:</b> Estrato geográfico.			

Fuente: Elaboración propia, en base a la información del ENAHG - INEI

### 3.13.2. Operacionalización de variables para el objetivo específico 2

Para este objetivo específico las variables se clasificaron de la siguiente manera (ver Tabla N° 6):

- Variable dependiente: Uso de internet.
- Variable independiente: Factores de uso de internet.

**Tabla 6**

*Resumen de las variables para el O.E.2.*

Variables	Definición conceptual	Definición operacional	Dimensiones	Indicadores	Ítems
<b>Variable 3:</b> <b>Uso de internet</b> (Variable dependiente)	Es la disposición o decisión de un jefe de hogar para hacer el uso de internet o no, en cualquier lugar.	Esta variable fue medida con el cuestionario de la Encuesta Nacional de Hogares (ENAHO), a través del módulo 300.	Uso de internet	1=Usa el internet 0=No usa el internet	<b>p314a:</b> En el mes anterior, ¿Ud. hizo uso del servicio de internet?
			Edad	Número de años cumplidos	<b>p208a:</b> ¿Qué edad tiene en años cumplidos? (en años).
			Genero	1=Hombre 0=Mujer	<b>p207:</b> Sexo.
			Estado civil	1=Casado 0=Soltero	<b>p209:</b> ¿Cuál es su estado civil o conyugal?
			Educación	1=Inicial 2=Primaria 3=Secundaria 4=Superior tecnológica 5=Superior universitaria 6=Maestría/doctorado	<b>p301a:</b> ¿Cuál es el último año o grado de estudios y nivel que aprobó? – nivel.
<b>Variable 4:</b> <b>Factores de uso de internet</b> (Variable independiente)	Los factores de uso de internet son las covariables que explican el simple hecho de que un jefe de hogar decida o no usar el internet, en cualquier lugar.	Esta variable fue medida con el cuestionario de la Encuesta Nacional de Hogares (ENAHO), a través de los módulos 200, 300, 612 y 34.	Miembros	Número de miembros	<b>mieperho:</b> Total de miembros del hogar.
			Niños	1=Si 0=No	<b>p208a:</b> ¿Qué edad tiene en años cumplidos?
			Estudiantes	1=Si 0=No	<b>p303:</b> El año pasado (...), ¿estuvo matriculado en algún centro o programa de educación básica o superior?
					<b>p306:</b> Este año, ¿está matriculado en algún centro o programa de educación básica o superior?
			Celular	1=Si 0=No	<b>p1142:</b> su hogar tiene: ¿teléfono celular?
			Computadora	1=Si 0=No	<b>p612n:</b> Equipamiento del hogar. <b>p612:</b> ¿Su hogar tiene?
			Región geográfica	1=Costa 2=Sierra 3=Selva	<b>dominio:</b> Dominio geográfico.
			Zona de residencia	1=Zona urbana 0=Zona rural	<b>estrato:</b> Estrato geográfico.

Fuente: Elaboración propia, en base a la información del ENAHO - INEI

### 3.13.3. Operacionalización de variables para el objetivo específico 3

Para este objetivo específico las variables se clasificaron de la siguiente manera (ver Tabla N° 7):

- Variable de resultado: Empleo.
- Variable de tratamiento: Uso de internet.
- Variable de control: Características observables.

**Tabla 7**

*Resumen de las variables para el O.E.3.*

Variables	Definición conceptual	Definición operacional	Dimensiones	Indicadores	Ítems
<b>Variable Ingreso</b> (Variable de resultado)	El ingreso es el dinero que recibe un agente económico debido a su actividad económica.	Esta variable fue medida con el cuestionario de la Encuesta Nacional de Hogares (ENAHG), a través de los módulos 500 y 34.	Ingreso	Ingreso neto per cápita	<b>inghog2d:</b> Ingreso neto total. <b>ocu500:</b> Indicador de la pea.
<b>Variable 6: Uso de internet</b> (Variable de tratamiento)	Es la disposición o decisión de un jefe de hogar en condición de PEA para hacer el uso de internet o no, desde cualquier lugar.	Esta variable fue medida con el cuestionario de la Encuesta Nacional de Hogares (ENAHG), a través del módulo 300.	Uso de internet	1=Usa el internet 0=No usa el internet	<b>p314a:</b> En el mes anterior, ¿Ud. hizo uso del servicio de internet?
<b>Variable 7: Características observables</b> (Variable de control)	Es la condición que determina el uso o no uso del internet, por parte del jefe de hogar, desde cualquier lugar.	Esta variable fue medida con el cuestionario de la Encuesta Nacional de Hogares (ENAHG), a través de los módulos 200, 300, 612 y 34.	Edad	Número de años cumplidos	<b>p208a:</b> ¿Qué edad tiene en años cumplidos? (en años).
			Genero	1=Hombre 0=Mujer	<b>p207:</b> Sexo.
			Estado civil	1=Casado 0=Soltero	<b>p209:</b> ¿Cuál es su estado civil o conyugal?
			Educación	1=Inicial 2=Primaria 3=Secundaria 4=Superior tecnológica 5=Superior universitaria 6=Maestría/doctorado	<b>p301a:</b> ¿Cuál es el último año o grado de estudios y nivel que aprobó? – nivel.
			Miembros	Número de miembros	<b>mieperho:</b> Total de miembros del hogar.
			Niños	1=Si 0=No	<b>p208a:</b> ¿Qué edad tiene en años cumplidos?
			Estudiantes	1=Si 0=No	<b>p303:</b> El año pasado (...), ¿estuvo matriculado en



Variables	Definición conceptual	Definición operacional	Dimensiones	Indicadores	Ítems
					algún centro o programa de educación básica o superior? <b>p306:</b> Este año, ¿está matriculado en algún centro o programa de educación básica o superior?
			Celular	1=Si 0=No	<b>p1142:</b> su hogar tiene: ¿teléfono celular?
			Computadora	1=Si 0=No	<b>p612n:</b> Equipamiento del hogar. <b>p612:</b> ¿Su hogar tiene?
			Región geográfica	1=Costa 2=Sierra 3=Selva	<b>dominio:</b> Dominio geográfico.
			Zona de residencia	1=Zona urbana 0=Zona rural	<b>estrato:</b> Estrato geográfico.

Fuente: Elaboración propia, en base a la información del ENAHO - INEI

## CAPÍTULO IV

### RESULTADOS Y DISCUSIÓN

#### 4.1. RESULTADOS

##### 4.1.1. Presentación de las estadísticas descriptivas

##### 4.1.1.1. Análisis descriptivo de las variables explicativas

La Tabla N° 8 muestra los estadísticos descriptivos de las variables explicativas que se usaron para alcanzar los objetivos específicos de esta investigación. Dichos estadísticos fueron calculados en base a la información extraída de la ENAHO del año 2020, a nivel de todo el Perú.

**Tabla 8**

*Resumen de las estadísticas descriptivas*

<b>Variable</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Min</b>	<b>Max</b>	<b>Obs</b>
Edad	50.3237	13.8208	16	98	26,663
Genero	0.7404	0.4384	0	1	26,663
Est_civil	0.6691	0.4705	0	1	26,663
Sin_nivel	0.0489	0.2158	0	1	26,663
Primaria	0.3417	0.4743	0	1	26,663
Secundaria	0.3731	0.4836	0	1	26,663
Sup_no_univ	0.1155	0.3196	0	1	26,663
Sup_univ	0.0992	0.2990	0	1	26,663
Maestria	0.0215	0.1449	0	1	26,663
Miembros	3.5263	1.8047	1	18	26,663
Niños	0.3006	0.4585	0	1	26,663
Estudiantes	0.5980	0.4903	0	1	26,663
Celular	0.9363	0.2443	0	1	26,663
Computadora	0.2878	0.4528	0	1	26,663
Material	0.7154	0.4512	0	1	26,663
Habitaciones	3.2499	1.6124	1	15	26,663
Electricidad	0.9242	0.2647	0	1	26,663
Costa	0.3628	0.4808	0	1	26,663
Sierra	0.4140	0.4926	0	1	26,663
Selva	0.2232	0.4164	0	1	26,663
Urbano	0.5598	0.4964	0	1	26,663

Fuente: Elaboración propia



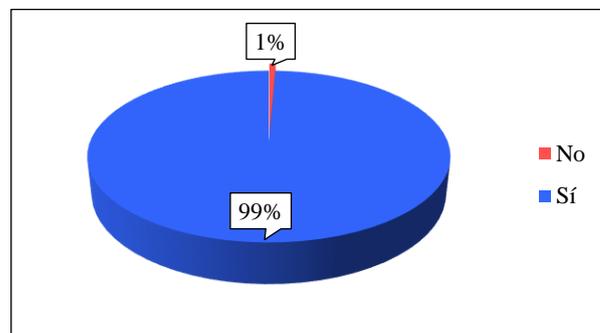
Como se puede observar en la Tabla N° 8, en el año 2020, a nivel de todo el Perú, la edad promedio de los jefes de hogar fue de 50 años aproximadamente; respecto al género, el 74% de los jefes de hogar fueron varones y el 26% fueron mujeres aproximadamente; en relación al estado civil, el 67% de los jefes de hogar fueron casado y el 33% fueron no casados; acerca de la educación, el 5% de los jefes de hogar no tenían algún nivel de educativo, el 34% tenían primaria, el 37% tenían secundaria, el 12% tenían superior no universitaria, el 10% tenían superior universitaria y solo el 2% tenían maestría y/o doctorado; respecto a los miembros, los hogares estuvieron conformados en promedio por 4 miembros aproximadamente; referente a los niños, el 30% de los hogares estuvieron conformados por al menos un niño; en cuanto a los estudiantes, el 60% de los hogares estuvieron conformados por al menos un estudiante; respecto al celular, el 94% de los jefes de hogar al menos disponían de un celular; respecto a la computadora, el 29% de los jefes de hogar al menos disponían de una computadora y/o laptop; en relación al material, el 72% de los hogares fueron construidos con material precario y el 28% con material no precario; respecto a las habitaciones, los hogares tuvieron en promedio 4 habitaciones aproximadamente; acerca de la electricidad; el 92% de los hogares tenían la instalación de la energía eléctrica; referente a la región geográfica, el 36% de los hogares se ubicaban en la región costa, el 41% en la región sierra y el 22% en la región selva; y por último, en cuanto al urbano, el 56% de los hogares se ubicaban en la zona urbana y el 44% en la zona rural.

#### 4.1.1.2. Análisis descriptivo de la variable PEA

En la Figura N° 7 se evidencia que, en el año 2020, a nivel de todo el Perú, el 99% de los jefes de hogar pertenecían a la PEA, los cuales fueron conformado por los jefes de hogar que estuvieron trabajando y no estuvieron trabajando, pero estuvieron buscando un empleo, mientras que el 1% pertenecían a la no PEA, los cuales fueron conformado por los adultos mayores, niños, jubilados, discapacitados, etc. Cabe aclarar que esta variable no interviene directamente en el logro de los objetivos, sin embargo, se utilizó indirectamente para lograr el objetivo específico 3, ya que en este objetivo solo se consideró a aquellos jefes de hogar que pertenecieron a la PEA.

**Figura 7**

*Población económicamente activa (PEA)*



Fuente: Elaboración propia

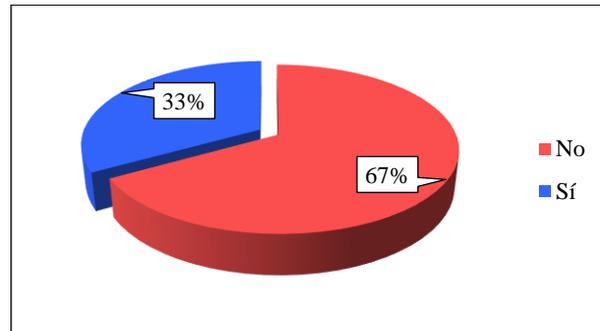
#### 4.1.1.3. Análisis descriptivo de la variable acceso a internet

En la Figura N° 8 se muestra que, en el año 2020, a nivel de todo el Perú, el 33% de los hogares tenían acceso a internet, mientras que el 67% no tenían acceso. Cabe aclarar que, el acceso a internet se refiere a la conexión de internet fijo desde una red pública hasta el interior del hogar.

Además, es preciso mencionar que esta variable se utilizó para alcanzar el objetivo específico 1, en donde tomó el rol de la variable dependiente.

**Figura 8**

*Conexión a internet en casa*



Fuente: Elaboración propia

Por otra parte, en la Tabla N° 9, se muestra que, en el año 2020, a nivel de departamentos, los hogares que se encontraron en Arequipa, Callao, Huánuco, Lima, Moquegua y Tacna tuvieron mayor acceso a internet, superiores al 50%, respecto a los demás departamentos.

**Tabla 9**

*Conexión a internet en casa, por departamentos*

Departamento	No	Sí	Total
Amazonas	74%	26%	100%
Ancash	72%	28%	100%
Apurímac	77%	23%	100%
Arequipa	49%	51%	100%
Ayacucho	83%	17%	100%
Cajamarca	86%	14%	100%
Callao	42%	58%	100%
Cusco	90%	10%	100%
Huancavelica	90%	10%	100%
Huánuco	78%	22%	100%
Ica	44%	56%	100%
Junín	74%	26%	100%
La Libertad	65%	35%	100%
Lambayeque	59%	41%	100%
Lima	46%	54%	100%

Departamento	No	Sí	Total
Loreto	74%	26%	100%
Madre de Dios	65%	35%	100%
Moquegua	46%	54%	100%
Pasco	71%	29%	100%
Piura	65%	35%	100%
Puno	82%	18%	100%
San Martín	68%	32%	100%
Tacna	46%	54%	100%
Tumbes	61%	39%	100%
Ucayali	76%	24%	100%
<b>Total</b>	<b>67%</b>	<b>33%</b>	<b>100%</b>

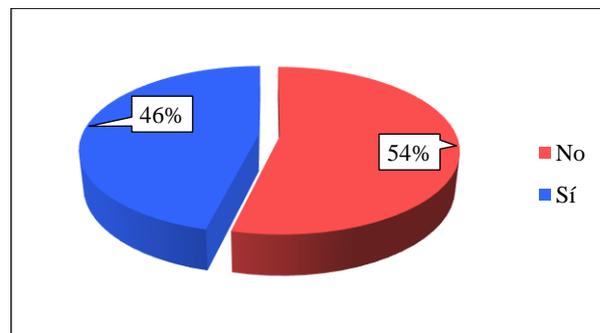
Fuente: Elaboración propia

#### 4.1.1.4. Análisis descriptivo de la variable uso de internet

En la Figura N° 9 se muestra que, en el año 2020, a nivel de todo el Perú, el 46% de los jefes de hogar usaron el internet, mientras el 54% no usaron. Esta variable se utilizó para alcanzar el objetivo específico 2, en donde tomó el rol de la variable dependiente. Asimismo, se utilizó para alcanzar el objetivo específico 3, en donde tomó el rol de la variable de tratamiento.

**Figura 9**

*Uso del servicio de internet*



Fuente: Elaboración propia

Por otra parte, en la Tabla N° 10, se muestra que, en el año 2020, a nivel de departamentos, los jefes de hogar que se residían en Arequipa,

Callao, Ica, La Libertad, Lambayeque, Lima, Madre de Dios, Moquegua, Piura, Tacna y Tumbes hicieron mayor uso de internet, mayores al 50%, respecto a los demás departamentos.

**Tabla 10**

*Uso del servicio de internet, por departamentos*

<b>Departamento</b>	<b>No</b>	<b>Sí</b>	<b>Total</b>
Amazonas	70%	30%	100%
Ancash	58%	42%	100%
Apurímac	73%	27%	100%
Arequipa	39%	61%	100%
Ayacucho	75%	25%	100%
Cajamarca	70%	30%	100%
Callao	22%	78%	100%
Cusco	77%	23%	100%
Huancavelica	76%	24%	100%
Huánuco	68%	32%	100%
Ica	30%	70%	100%
Junín	67%	33%	100%
La Libertad	44%	56%	100%
Lambayeque	47%	53%	100%
Lima	28%	72%	100%
Loreto	66%	34%	100%
Madre de Dios	43%	57%	100%
Moquegua	39%	61%	100%
Pasco	64%	36%	100%
Piura	48%	52%	100%
Puno	67%	33%	100%
San Martín	57%	43%	100%
Tacna	35%	65%	100%
Tumbes	38%	62%	100%
Ucayali	54%	46%	100%
<b>Total</b>	<b>54%</b>	<b>46%</b>	<b>100%</b>

Fuente: Elaboración propia

#### **4.1.1.5. Análisis descriptivo de la variable ingreso per cápita**

En la Tabla N° 11 se observa que, en el año 2020, a nivel de todo el Perú, el promedio del ingreso per cápita de los jefes de hogar fue de

799.90 nuevos soles, mientras que el promedio del logaritmo natural del ingreso per cápita de estas mismas personas fue de 6.31 nuevos soles.

**Tabla 11**

*Ingreso neto per cápita, en términos de meses*

<b>Estadísticos</b>	<b>Ingreso per cápita</b>	<b>Logaritmo del ingreso per cápita</b>
Media	799.90	6.31
Percentil (50%)	533.98	6.28
Máximo	29434.04	10.29
Mínimo	34.12	3.53
Desviación estándar	938.09	0.83
Varianza	880014.30	0.69
Asimetría	5.97	0.25
Curtosis	80.86	3.09

Fuente: Elaboración propia

Es importante aclarar que la variable ingreso per cápita en su estado natural no tiene una distribución normal, debido a que el valor del estadístico de la asimetría es 5.97 mayor a 0, lo cual indica que la distribución tiene una asimetría positiva y hay un sesgo hacia el lado derecho, es decir, la cola de la distribución se alarga a la derecha para valores superiores a la media. Además, el valor del estadístico de la curtosis es 80.86 mayor a 3, lo cual indica que la distribución no es normal, sino es una distribución leptocúrtica. (Ver Tabla N° 11).

Por esta razón, el ingreso per cápita se ha trabajado en base al logaritmo natural, a fin de suavizar la distribución de esta variable. En efecto, como se puede observar la Tabla N° 11, se muestra que el estadístico de la asimetría tiene un valor de 0.25 muy cercano a 0 y el estadístico de la curtosis tiene un valor de 3.09 muy cercano a 3, lo cual indica que la distribución de frecuencias de la variable logaritmo natural

del ingreso per cápita tiene una distribución simétrica y mesocúrtica, respectivamente, es decir, una distribución normal. Además, la media con un valor de 6.31 y la mediana con un valor de 6.28 son casi similares, lo que también evidencia que el logaritmo natural de esta variable tiene una distribución normal. En consecuencia, esta variable se utilizó para alcanzar el objetivo específico 3, el cual se consideró como la variable de resultado.

#### 4.1.2. Presentación de los resultados para el objetivo específico 1

En la sección de materiales y métodos se presentó el modelo econométrico que se utilizó para alcanzar el objetivo específico 1, el cual consta de una variable dependiente y 11 variables independientes o también conocido como covariables o variables explicativas, tal como se muestra a continuación:

$$\begin{aligned} Pr(\text{Acceso a internet} = 1) & \\ &= F[\alpha_0 + \alpha_1(\text{Material}) + \alpha_2(\text{Habitaciones}) \\ &+ \alpha_3(\text{Electricidad}) + \alpha_4(\text{Miembros}) + \alpha_5(\text{Niños}) \\ &+ \alpha_6(\text{Estudiantes}) + \alpha_7(\text{Celular}) + \alpha_8(\text{Computadora}) \\ &+ \alpha_9(\text{Sierra}) + \alpha_{10}(\text{Selva}) + \alpha_{11}(\text{Urbano}) + \mu_i] \end{aligned}$$

... Ecuación 48

##### 4.1.2.1. Análisis del modelo inicial

La Tabla N° 12 muestra las regresiones estimadas de los modelos Logit y Probit, en su versión inicial.

**Tabla 12***Coefficientes de los modelos iniciales para el O.E.1.*

<b>Covariables</b>	<b>Modelo Logit Acceso a internet (Coeficiente)</b>	<b>Modelo Probit Acceso a internet (Coeficiente)</b>
Material	-0.6888*** (-17.12)	-0.4064*** (-17.32)
Habitaciones	0.1799*** (15.46)	0.0993*** (15.13)
Electricidad	0.5843*** (5.01)	0.283*** (4.94)
Miembros	-0.0183 (-1.33)	-0.0099 (-1.28)
Niños	0.0194 (0.45)	0.0131 (0.53)
Estudiantes	0.2943*** (6.77)	0.1665*** (6.8)
Celular	2.7905*** (10.83)	1.4745*** (11.34)
Computadora	2.1239*** (57.04)	1.2695*** (58.36)
Sierra	-0.6559*** (-15.56)	-0.3705*** (-15.55)
Selva	-0.2703*** (-5.65)	-0.161*** (-5.93)
Urbano	1.0085*** (23.02)	0.5634*** (23.17)

-Los números entre paréntesis son los estadísticos z.

-Los asteriscos indican el nivel de significancia: \*\*\* al 1% y \*\* al 5%.

Fuente: Elaboración propia

Como se puede observar en la Tabla N° 12, en ambos modelos las covariables miembros y niños no son estadísticamente significativos, debido a que los valores del z-estadístico de los parámetros estimados en el modelo Logit son -1.33 y 0.45, respectivamente, mientras que en el modelo Probit son -1.28 y 0.53, respectivamente, menores a 2 en valor absoluto en ambos modelos, el cual indica que estos valores caen en una zona de no rechazo de la distribución normal o z, por lo tanto, no es posible rechazar la hipótesis nula ( $\alpha_i = 0$ ), mostrando de esta manera que el

coeficiente estimado de estas covariables son igual a 0. Además, en ambas covariables sucede que las probabilidades de rechazar la hipótesis nula de sus coeficientes estimados son mayores al 5%, por lo tanto, esto muestra que tampoco se puede rechazar la hipótesis nula (ver Tabla B1 y Tabla B4).

En consecuencia, se procede a eliminar estas covariables a fin de maximizar las predicciones correctas de las probabilidades y tener un buen ajuste en ambos modelos.

#### 4.1.2.2. Análisis del modelo final

La Tabla N° 13 muestra las regresiones estimadas de los modelos Logit y Probit, en su versión final.

**Tabla 13**

*Coefficientes y medidas de ajuste de los modelos finales para el O.E.1.*

<b>Covariables y medidas de ajuste</b>	<b>Modelo Logit Acceso a internet (Coeficiente)</b>	<b>Modelo Probit Acceso a internet (Coeficiente)</b>
Material	-0.6925*** (-17.26)	-0.4084*** (-17.45)
Habitaciones	0.1760*** (15.66)	0.0972*** (15.32)
Electricidad	0.5904*** (5.07)	0.2858*** (4.99)
Estudiantes	0.2624*** (7.25)	0.1498*** (7.35)
Celular	2.7839*** (10.81)	1.4712*** (11.32)
Computadora	2.1230*** (57.21)	1.2689*** (58.51)
Sierra	-0.6533*** (-15.51)	-0.3691*** (-15.52)
Selva	-0.2695*** (-5.64)	-0.1606*** (-5.91)
Urbano	1.0105*** (23.08)	0.5645*** (23.23)

Covariables y medidas de ajuste	Modelo Logit Acceso a internet (Coeficiente)	Modelo Probit Acceso a internet (Coeficiente)
Modelo completo con Log-Lik	-10769.010	-10763.849
LR $\chi^2$ (9)	12332.659	12342.980
Prob > $\chi^2$	0.000	0.000
$R^2$ de McFadden	0.364	0.364
$R^2$ ajustado de McFadden	0.364	0.364
$R^2$ de ML (Cox-Snell)	0.370	0.371
$R^2$ de Cragg-Uhler (Nagelkerke)	0.515	0.515
$R^2$ de McKelvey y Zavoina	0.549	0.565
$R^2$ de Efron	0.424	0.424
$R^2$ de recuento	0.825	0.825
$R^2$ ajustado de recuento	0.472	0.472
AIC	0.809	0.808
BIC	-250083.559	-250093.880
Área bajo la curva ROC	0.877	0.877

-Los números entre paréntesis son los estadísticos z.  
-Los asteriscos indican el nivel de significancia: \*\*\* al 1% y \*\* al 5%.

Fuente: Elaboración propia

Como se puede observar en la Tabla N° 13, una vez excluidos las covariables miembros y niños, los coeficientes estimados para los covariables restantes, en ambos modelos resultaron ser significativos al 1%. Es decir, dado que sus valores del z-estadístico son mayores a 2 en valor absoluto, lo cual muestra que estos valores caen en una zona de rechazo de la distribución normal o z, entonces es factible rechazar la hipótesis nula ( $\alpha_i = 0$ ), indicando de esta manera que los coeficientes estimados son estadísticamente distintos de 0. Además, en todas las covariables se evidencia que sus probabilidades de rechazar a la hipótesis nula son menores al 1%, y por ende también se debe rechazar la hipótesis nula (ver Tabla B2 y Tabla B5). En ese contexto, se puede afirmar que en ambos modelos existe significancia individual en todos sus parámetros estimados.



También, se puede afirmar que los parámetros estimados en conjunto son significativos en ambos modelos (significancia global), ya que la probabilidad de rechazar la hipótesis nula es 0%, menor al 5%, por lo tanto, es posible rechazar la hipótesis nula ( $\alpha_1 = \dots = \alpha_k = 0$ ) (ver Tabla B2 y Tabla B5). Además, el valor del  $\chi^2$ -estadístico cae en una zona de rechazo de la distribución  $\chi^2$ , lo cual muestra que también existe significancia conjunta en ambos modelos.

Por otra parte, para elegir el mejor modelo se debe tomar en cuenta la función de verosimilitud logarítmica, el estadístico LR o razón de verosimilitud, las medidas de bondad de ajuste, los estadísticos de Akaike y Schwarz de pérdida de información y el área bajo la curva ROC. Por lo tanto, en la Tabla N° 13 se observa que el mejor modelo es el modelo Probit, debido a que este tiene mayor verosimilitud, mayor significancia conjunta o razón de verosimilitud, mayores  $R^2$  y valores bajos en los criterios de información. Cabe aclarar que el área bajo la curva ROC para ambos modelos son lo mismo, así que este indicador no es indispensable para la elección del mejor modelo.

Además, respecto a los parámetros estimados en una regresión no lineal se debe tener en cuenta que estos no son interpretables directamente, sin embargo, se puede saber el efecto que ocasiona una variable explicativa en la variable dependiente mediante los signos. En ese sentido, se observa que las covariables como: material, sierra y selva afectan negativamente a la probabilidad de que un hogar pueda tener acceso a internet disponible, mientras que las covariables como: habitaciones, electricidad, estudiantes,

celular, computadora y urbano afectan positivamente a la probabilidad de que un hogar tenga acceso a internet disponible.

#### 4.1.2.3. Estimación de los efectos marginales

Una vez elegido el mejor modelo es factible estimar los efectos marginales de cada covariable. En ese sentido, la Tabla N° 14 muestra los efectos marginales de cada variable explicativa sobre la probabilidad de acceso a internet disponible, evaluado en sus valores medios.

**Tabla 14**

*Efectos marginales del modelo elegido para el O.E.1.*

<b>Covariables</b>	<b>Modelo Probit Acceso a internet (Impacto marginal)</b>
Material	-0.1381*** (-16.67)
Habitaciones	0.0313*** (-15.24)
Electricidad	0.0842*** (-5.56)
Estudiantes	0.0477*** (-7.43)
Celular	0.2662*** (35.81)
Computadora	0.4444*** (58.06)
Sierra	-0.1188*** (-15.5)
Selva	-0.0551*** (-5.98)
Urbano	0.1764*** (24.3)

-Los números entre paréntesis son los estadísticos z.

-Los asteriscos indican el nivel de significancia: \*\*\* al 1% y \*\* al 5%.

Fuente: Elaboración propia

La interpretación de los efectos marginales derivados de los coeficientes de la regresión del modelo Probit, está basado en los cambios



que ocasionaron las variables explicativas en la probabilidad de que un hogar del Perú tuvo acceso a internet disponible en el año 2020, tal como se muestra a continuación:

- Un hogar que fue construido con material rustico tenía un 13.81% menos probabilidad de tener conexión a internet disponible, en comparación a un hogar que fue construido con material no rustico. Por lo tanto, los hogares que fueron construidos con material rustico (adobe, piedra, tierra, barro, madera, triplay, calamina, estera, paja, etc.) tuvieron menores posibilidades de tener acceso a internet, mientras que los hogares que fueron construidos con material no rustico (ladrillo, bloqueta, parquet, cemento, laminas asfálticas, losetas, concreto armado, tejas, etc.) tuvieron mayores posibilidades de tener acceso a internet. Esto podría deberse a que los hogares construidos con material rustico se encuentren en zonas donde no existe la infraestructura de las redes de conexión a internet, o simplemente existen estas redes, pero el hogar está catalogado como pobre y sus miembros no tienen suficientes recursos para hacer conectar el internet hasta el interior del hogar.
- Por cada habitación adicional que tenía un hogar, su probabilidad de tener conexión a internet disponible se incrementaba en un 3.13%. Por consiguiente, a mayor número de habitaciones, los hogares tuvieron mayores posibilidades de tener acceso a internet. Este acontecimiento se podría haberse dado porque a mayor número de miembros, mayor es la necesidad de tener conexión a internet en el hogar o internet fijo, por el simple hecho de que



teniendo internet móvil para cada miembro, mayores son los gastos, en cambio, si el hogar tendría internet fijo, entonces los miembros pueden compartir internet mediante la conexión de wifi y con eso estarían disminuyendo sus gastos.

- Un hogar que contaba con energía eléctrica tenía un 8.42% más probabilidad de tener conexión a internet disponible, respecto a un hogar que no contaba con energía eléctrica. Por lo tanto, los hogares que disponían de electricidad tuvieron mayores posibilidades de tener acceso a internet, mientras que los hogares que no disponían de electricidad tuvieron menores posibilidades de tener acceso a internet. Esto concuerda con la realidad, debido a que si no hay energía eléctrica no habrá infraestructura de redes de conexión a internet, por lo tanto, la disponibilidad de electricidad en el hogar es de suma importancia para que sus miembros decidan a hacerse instalar el internet desde una red pública, asimismo, para usar los dispositivos electrónicos.
- Si en un hogar existía estudiantes, su probabilidad de tener conexión a internet disponible aumentaba en 4.77%, respecto a un hogar en donde no existía estudiantes. En consecuencia, en los hogares donde había estudiantes tenían mayores posibilidades de tener acceso a internet, en cambio, en los hogares donde no había ningún estudiante tenían menores posibilidades de tener acceso a internet. Esto podría haber sucedido porque los estudiantes necesitan de una conexión a internet fijo para realizar sus tareas y trabajos, comunicarse con sus compañeros, realizar reuniones



virtuales, investigar, etc., por lo tanto, esto conlleva a los miembros de un hogar a tomar la decisión de hacerse instalar el internet fijo desde una red pública.

- Si en un hogar había algún miembro que tenía teléfono móvil, su probabilidad de que tenga conexión a internet disponible se incrementaba en 26.62%, en comparación a un hogar en donde había ningún miembro que tenía este aparato electrónico. En este contexto, se puede afirmar que, en aquellos hogares en donde había algún miembro que tuviese celular tenían más posibilidades de tener acceso a internet, mientras que en los hogares en donde no había ningún miembro que tuviese este aparato electrónico tenían menos posibilidades de tener acceso a internet. Obviamente, si en un hogar existe algún miembro que tenga celular querrá hacerse instalar internet fijo desde una red pública, ya sea para fines de trabajo, educativos o distracción.
- Si en un hogar existía algún miembro que tenía computadora y/o laptop, su probabilidad de que tenga conexión a internet disponible se incrementaba en 44.44%, respecto a un hogar en donde no existía algún miembro que tenía estos artefactos. Por consiguiente, se puede aseverar que, en aquellos hogares en donde había algún miembro que tuviese computadora y/o laptop tenían más posibilidades de tener acceso a internet, mientras que en los hogares en donde no había ningún miembro que tuviese estos artefactos tenían menos posibilidades de tener acceso a internet. Obviamente, si en un hogar existe algún miembro que tenga



computadora y/o laptop querrá hacerse instalar internet fijo desde una red pública, ya sea para fines de trabajo, educativos o distracción.

- Un hogar que se ubicaba en la sierra o selva tenía sucesivamente un 11.88% o 5.51% menos probabilidad de tener conexión a internet disponible, en comparación a un hogar que se ubicaba en la costa. En consecuencia, los hogares que se ubicaban en la costa tuvieron mayores posibilidades de tener acceso a internet, respecto a los hogares que se ubicaban en la selva o sierra. Asimismo, los hogares que se ubicaban en la selva tuvieron mayores posibilidades de tener acceso a internet, en comparación a los hogares que se ubicaban en la sierra. Este acontecimiento se podría haberse dado porque quizá la infraestructura de las redes de conexión a internet fijo no haya llegado a todos los rincones de la sierra o selva, o simplemente si hayan llegado, pero el hogar está catalogado como pobre y sus miembros no tienen suficientes recursos para hacerse conectar el internet hasta el interior del hogar. (Ver Tabla B8).
- Un hogar que se ubicaba en la zona urbana tenía un 17.64% más probabilidad de tener conexión a internet disponible, en comparación a los hogares que se ubicaban en la zona rural. Por lo tanto, los hogares de la zona urbana tuvieron mayores posibilidades de tener acceso a internet, mientras que los hogares de la zona rural tuvieron menores posibilidades. Esto indica que en la zona urbana la infraestructura de las redes de conexión a internet está distribuida casi en todo su territorio.

#### 4.1.3. Presentación de los resultados para el objetivo específico 2

En la sección de materiales y métodos se presentó el modelo econométrico que se utilizó para alcanzar el objetivo específico 2, el cual consta de una variable dependiente y 16 variables independientes o variables explicativas, tal como se muestra a continuación:

$$\begin{aligned} Pr(\text{Uso de internet} = 1) &= F[\beta_0 + \beta_1(\text{Edad}) + \beta_2(\text{Genero}) + \beta_3(\text{Est_civil}) \\ &+ \beta_4(\text{Primaria}) + \beta_5(\text{Secundaria}) + \beta_6(\text{Sup_no_univ}) \\ &+ \beta_7(\text{Sup_univ}) + \beta_8(\text{Maestria}) + \beta_9(\text{Miembros}) \\ &+ \beta_{10}(\text{Niños}) + \beta_{11}(\text{Estudiantes}) + \beta_{12}(\text{Celular}) \\ &+ \beta_{13}(\text{Computadora}) + \beta_{14}(\text{Sierra}) + \beta_{15}(\text{Selva}) \\ &+ \beta_{16}(\text{Urbano}) + \varepsilon_i] \end{aligned}$$

... Ecuación 49

##### 4.1.2.1. Análisis del modelo inicial

La Tabla N°15 muestra las regresiones estimadas de los modelos Logit y Probit, en su versión inicial.

**Tabla 15**

*Coefficientes de los modelos iniciales para el O.E.2.*

Covariables	Modelo Logit Uso de internet (Coeficiente)	Modelo Probit Uso de internet (Coeficiente)
Edad	-0.0749*** (-42.06)	-0.0425*** (-43.06)
Genero	0.2126*** (4.03)	0.1178*** (3.91)
Est_civil	-0.2593*** (-4.85)	-0.1395*** (-4.58)

Covariables	Modelo Logit Uso de internet (Coeficiente)	Modelo Probit Uso de internet (Coeficiente)
Primaria	1.5234*** (6.97)	0.7411*** (7.21)
Secundaria	2.6669*** (12.25)	1.4162*** (13.82)
Sup_no_univ	3.9323*** (17.6)	2.1325*** (20.1)
Sup_univ	4.6697*** (20.31)	2.5374*** (23.14)
Maestria	6.9075*** (13.7)	3.4898*** (17.23)
Miembros	-0.1182*** (-7.89)	-0.0671*** (-7.89)
Niños	0.0700 (1.52)	0.0374 (1.41)
Estudiantes	0.1242*** (2.61)	0.0711*** (2.62)
Celular	2.6556*** (15.9)	1.4627*** (16.78)
Computadora	1.0289*** (23.17)	0.5957*** (23.43)
Sierra	-1.0520*** (-24.11)	-0.5993*** (-24.11)
Selva	-0.8861*** (-18.52)	-0.5039*** (-18.41)
Urbano	0.9177*** (22.47)	0.5292*** (22.48)

-Los números entre paréntesis son los estadísticos z.

-Los asteriscos indican el nivel de significancia: \*\*\* al 1% y \*\* al 5%.

Fuente: Elaboración propia

Como se puede observar en la Tabla N° 15, la covariable niños no es estadísticamente significativo en ambos modelos, debido a que los valores del z-estadístico de los parámetros estimados en el modelo Logit es igual a 1.52 y en el modelo Probit es igual a 1.41, ambos menores a 2 en valor absoluto, lo cual muestra que estos valores caen en una zona de no rechazo de la distribución normal tipificada o z, por lo tanto, no es posible rechazar la hipótesis nula ( $\beta = 0$ ), evidenciando de esta manera que el coeficiente estimado de esta covariable es estadísticamente igual a

0. Además, las probabilidades de rechazar la hipótesis nula de los coeficientes estimados para ambos modelos son mayores al 5%, lo cual indica que tampoco es posible rechazar la hipótesis nula (ver Tabla C1 y Tabla C4).

Por consiguiente, se procede a eliminar estas covariables a fin de maximizar las predicciones correctas de las probabilidades y tener un buen ajuste en ambos modelos.

#### 4.1.2.2. Análisis del modelo final

La Tabla N° 16 muestra las regresiones estimadas de los modelos Logit y Probit, en su versión final.

**Tabla 16**

*Coefficientes y medidas de ajuste de los modelos finales para el O.E.2.*

Covariables y medidas de ajuste	Modelo Logit Uso de internet (Coeficiente)	Modelo Probit Uso de internet (Coeficiente)
Edad	-0.0758*** (-44.78)	-0.0430*** (-45.96)
Genero	0.2106*** (3.99)	0.1167*** (3.88)
Est_civil	-0.2601*** (-4.86)	-0.1397*** (-4.59)
Primaria	1.5237*** (6.97)	0.7414*** (7.2)
Secundaria	2.6693*** (12.25)	1.4176*** (13.82)
Sup_no_univ	3.9355*** (17.6)	2.1345*** (20.1)
Sup_univ	4.6745*** (20.33)	2.5400*** (23.15)
Maestria	6.9132*** (13.71)	3.4926*** (17.23)
Miembros	-0.1076*** (-8.12)	-0.0615*** (-8.17)
Estudiantes	0.1114** (2.38)	0.0643** (2.41)
Celular	2.6516***	1.4610***

Covariables y medidas de ajuste	Modelo Logit Uso de internet (Coeficiente)	Modelo Probit Uso de internet (Coeficiente)
	(15.88)	(16.76)
Computadora	1.0222***	0.5921***
	(23.14)	(23.41)
Sierra	-1.0522***	-0.5994***
	(-24.11)	(-24.11)
Selva	-0.8841***	-0.5028***
	(-18.49)	(-18.38)
Urbano	0.9165***	0.5288***
	(22.45)	(22.47)
Modelo completo con Log-Lik	-10220.897	-10223.039
LR $\chi^2(15)$	16364.584	16360.301
Prob > $\chi^2$	0.000	0.000
$R^2$ de McFadden	0.445	0.444
$R^2$ ajustado de McFadden	0.444	0.444
$R^2$ de ML (Cox-Snell)	0.459	0.459
$R^2$ de Cragg-Uhler (Nagelkerke)	0.613	0.613
$R^2$ de McKelvey y Zavoina	0.696	0.704
$R^2$ de Efron	0.504	0.504
$R^2$ de recuento	0.821	0.821
$R^2$ ajustado de recuento	0.612	0.613
AIC	0.768	0.768
BIC	-251118.638	-251114.355
Área bajo la curva ROC	0.907	0.907

-Los números entre paréntesis son los estadísticos z.

-Los asteriscos indican el nivel de significancia: \*\*\* al 1% y \*\* al 5%.

Fuente: Elaboración propia

Como se observa en la Tabla N°16, una vez extraído la covariable niños, en ambos modelos se verifica que existe una significancia individual a nivel de todos los coeficientes estimados para cada covariable. Es decir, los valores del z-estadístico calculado para ambos modelos son mayores a 2 en valor absoluto, lo cual muestra que estos valores caen en una zona de rechazo de la distribución normal tipificada o z, por lo tanto, es factible rechazar la hipótesis nula ( $\beta_i = 0$ ), indicando de esta manera que los coeficientes estimados son estadísticamente diferentes de 0. Además, en todas las covariables para ambos modelos se evidencia que las

probabilidades de rechazar la hipótesis nula de sus parámetros estimados son menores al 1%, con excepción de la covariable estudiantes que es menor al 5%, y por ende también es factible rechazar la hipótesis nula (ver Tabla C2 y Tabla C5). Por lo tanto, esto indica que todas las variables explicativas son estadísticamente significativas al 1%, con excepción de la covariable estudiantes, que es significativo al 5%.

También, se puede afirmar que en ambos modelos los parámetros estimados en conjunto son significativos (significancia global), ya que la probabilidad de rechazar la hipótesis nula es menor al 5%, por lo tanto, es posible rechazar la hipótesis nula ( $\beta_1 = \dots = \beta_k = 0$ ) (ver Tabla C2 y Tabla C5). Además, el valor del  $\chi^2$ -estadístico cae en una zona de rechazo de la distribución  $\chi^2$ , lo cual muestra que también existe significancia conjunta en ambos modelos.

Por otra parte, para elegir el mejor modelo se debe tomar en cuenta la función de verosimilitud logarítmica, el estadístico LR o razón de verosimilitud, las medidas de bondad de ajuste, los estadísticos de Akaike y Schwarz de pérdida de información y el área bajo la curva ROC. En ese sentido, en la Tabla N° 16 se observa que el mejor modelo es el modelo Logit, debido a que este tiene mayor verosimilitud, mayor significancia conjunta o razón de verosimilitud, mayor  $R^2$  de McFadden y un valor bajo en los criterios de información bayesiano (BIC). Cabe mencionar que el área bajo la curva ROC, el AIC y algunas medidas de ajuste para ambos modelos son lo mismo, así que este indicador es indiferente para la elección del mejor modelo.

Además, respecto a los parámetros estimados en una regresión no lineal se debe tener en cuenta que estos no son interpretables directamente, sin embargo, se puede conocer el efecto que ocasiona una variable explicativa a la variable dependiente mediante los signos. En ese sentido, se observa que las covariables como: edad, estado civil, miembros, sierra y selva afectan negativamente a la probabilidad de que un hogar pueda tener acceso a internet disponible, mientras que las covariables como: genero, primaria, secundaria, superior no universitaria, superior universitaria, maestría y doctorado, estudiantes, celular, computadora y urbano afectan positivamente a la probabilidad de que un hogar tenga acceso a internet disponible.

#### 4.1.2.3. Estimación de los efectos marginales

Una vez elegido el mejor modelo es posible estimar los efectos marginales de cada covariable. En ese sentido, la Tabla N° 17 muestra los efectos marginales de cada variable explicativa sobre la probabilidad de uso de internet, evaluado en sus valores medios.

**Tabla 17**

*Efectos marginales del modelo elegido para el O.E.2.*

<b>Covariables</b>	<b>Modelo Logit Uso de internet (Impacto marginal)</b>
Edad	-0.0183*** (-44.78)
Genero	0.0503*** (4.04)
Est_civil	-0.0632*** (-4.84)
Primaria	0.1431*** (12.51)



<b>Covariables</b>	<b>Modelo Logit Uso de internet (Impacto marginal)</b>
Secundaria	0.3790*** (31.57)
Sup_no_univ	0.6772*** (44.61)
Sup_univ	0.7983*** (55.92)
Maestria	0.9318*** (70.68)
Miembros	-0.0260*** (-8.1)
Estudiantes	0.0268** (2.39)
Celular	0.3945*** (39.3)
Computadora	0.2487*** (23.7)
Sierra	-0.2529*** (-25.02)
Selva	-0.2158*** (-19.19)
Urbano	0.2156*** (23.47)

-Los números entre paréntesis son los estadísticos z.

-Los asteriscos indican el nivel de significancia: \*\*\* al 1% y \*\* al 5%.

Fuente: Elaboración propia

La interpretación de los efectos marginales derivados de los coeficientes de la regresión del modelo Logit, está basado en los cambios que ocasionaron las variables explicativas en la probabilidad de que un jefe de hogar del Perú hizo el uso de internet en el año 2020, tal como se muestra a continuación:

- Por cada año adicional que tenía un jefe de hogar, su probabilidad de usar el internet disminuyó en un 1.83%. Por lo tanto, esto indica que a medida que los jefes de hogar tenían más edad sus posibilidades de usar el internet se redujeron, o dicho de otro modo,



los jefes de hogar más jóvenes tenían mayores posibilidades de usar el internet que los jefes de hogar más longevos.

- Un jefe de hogar hombre tenía un 5.03% más probabilidad de usar el internet, en comparación a un jefe de hogar mujer. Por consiguiente, si los jefes de hogar eran hombres tenían mayores posibilidades de hacer el uso de internet, respecto a los jefes de hogar si eran mujeres. Esto podría deberse a que los hombres tenían mayor disponibilidad de tiempo o mayor necesidad de hacer el uso de internet, ya sea por temas de trabajo o simplemente por distracción, en cambio, las mujeres lo más probable eran las que no disponían de mucho tiempo, ya que talvez se enfocaban mayormente en los deberes del hogar, y por ende no disponían de mucho tiempo para hacer el uso de internet.
- Un jefe de hogar que estaba casado tenía un 6.32% menos probabilidad de usar el internet, en comparación a un jefe de hogar que se encontraba en la situación de no sacado. En consecuencia, los jefes de hogar que tenían el estado civil de casado o conviviente, tuvieron menos posibilidades de usar el internet frente a aquellos jefes de hogar que estaban solteros, viudos, divorciados o separados. Este acontecimiento se podría haber dado porque los jefes de hogar considerados en la categoría de casados, tuvieron muchas responsabilidades, ya que al estar en esta categoría necesariamente tuvieron que salir a trabajar para sostener a los miembros de su familia y por ende tenían poco o nada de tiempo para hacer el uso de internet, en cambio, los jefes de hogar que



pertenecían a la categoría de no casados, tenían pocas responsabilidades y por ende mayor disponibilidad de tiempo para usar el internet.

- Un jefe de hogar que había logrado estudiar la educación primaria tenía un 14.31% más probabilidad de usar el internet; uno que había logrado estudiar la educación secundaria tenía un 37.90% más probabilidad de usar el internet; uno que había logrado estudiar la educación superior no universitaria tenía un 67.72% más probabilidad de usar el internet; uno que había logrado estudiar la educación superior universitaria tenía un 79.83% más probabilidad de usar el internet; y uno que había logrado estudiar la maestría y doctorado tenía un 93.18% más probabilidad de usar el internet, en comparación a un jefe de hogar que solo había logrado estudiar la educación inicial o no había logrado ningún nivel educativo. Por lo tanto, se puede evidenciar que a medida que los jefes de hogar habían logrado alcanzar un alto nivel educativo tenían mayores posibilidades de hacer el uso de internet frente a aquellos jefes de hogar que solo habían logrado estudiar el nivel inicial o más aun no habían logrado estudiar algún nivel educativo. (Ver Tabla C8).
- Por cada miembro adicional que tenía un jefe de hogar, su probabilidad de usar el internet se redujo en 2.60%. En efecto, a mayor número de miembros, un jefe de hogar tenía menores posibilidades de usar el internet, o dicho de otro modo, los jefes de hogar que tenían pocos miembros en su familia tenían mayores posibilidades de usar el internet, por el simple hecho de que a



menor tamaño de familia menor responsabilidad tienen y por ende disponen de mayor tiempo para hacer el uso de internet.

- Un jefe de hogar que era estudiante y/o tenía algún miembro que estaba estudiando tuvo un 2.68% más probabilidad de usar el internet, en comparación a un jefe de hogar que no tenía miembros de su familia estudiando. Por consiguiente, los jefes de hogar que eran estudiantes y/o tenían algún miembro de su familia que estaban estudiante (hijos, esposo (a), conviviente, etc.) tenían mayores posibilidades de hacer el uso de internet. Esto podría deberse a que los estudiantes siempre tienen tareas, trabajos, actividades, reuniones grupales, etc., y muchos de ellos se realizan con la ayuda del internet, por lo tanto, la necesidad de hacer el uso del internet es muy alto para este grupo de personas, en cambio, para el grupo de personas que no son estudiantes hacer el uso de internet es poco o nada importante, ya que en su mayoría solo usan el internet para fines de distracción.
- Un jefe de hogar que disponía de un teléfono móvil tenía un 39.45% más probabilidad de usar el internet, en comparación a un jefe de hogar que no disponía de este aparato electrónico. Por lo tanto, a nivel del Perú, los jefes de hogar que contaban con celular tenían mayores posibilidades de hacer el uso de internet, en relación a aquellos jefes de hogar que no tenían celular. Esto concuerda con la realidad, ya que, para hacer la búsqueda de información, navegar, interactuar, entretenerse, etc., mediante el



internet, una persona tiene que acudir a algún medio, en este caso a un celular.

- Un jefe de hogar que disponía de una computadora y/o laptop tenía un 24.87% más probabilidad de usar el internet, en comparación de un jefe de hogar que no disponía de estos artefactos. Por consiguiente, a nivel del Perú, los jefes de hogar que contaban con computadora y/o laptop tenían mayores posibilidades de hacer el uso de internet, en relación a aquellos jefes de hogar que no tenían estos artefactos. Esto concuerda con la realidad, ya que, para hacer la búsqueda de información, navegar, interactuar, entretenerse, etc., a través del internet, una persona tiene que acudir a algún medio, en este caso a una computadora y/o laptop.
- Un jefe de hogar que vivía en una región geográfica sierra tenía un 25.29% menos probabilidad de usar el internet, en comparación a un jefe de hogar que vivía en la costa. Por lo tanto, a nivel de todo el Perú, los jefes de hogar que vivían en sierra, tenían menores posibilidades de hacer el uso de internet, frente a aquellos jefes de hogar que vivían en la costa. Este acontecimiento se podría haberse dado porque quizá la infraestructura de las redes de conexión a internet no haya llegado a todos los rincones de sierra, o también por la escasez de recursos económicos que tienen los jefes de hogar para contratar un servicio de internet o simplemente por la falta de conocimiento en el manejo de las TIC. (Ver Tabla N° C9).
- Un jefe de hogar que vivía en una región geográfica selva tenía un 21.58% menos probabilidad de usar el internet, en comparación a



un jefe de hogar que vivía en la costa. Por consiguiente, a nivel de todo el Perú, los jefes de hogar que vivían en selva, tenían menores posibilidades de hacer el uso de internet, frente a aquellos jefes de hogar que vivían en la costa. Esto indica que en algunas zonas de la selva no existe infraestructura de las redes de conexión a internet, o simplemente existen, pero algunos jefes de hogar no tienen suficientes recursos para contratar un servicio de internet, los cuales les dificulta para hacer el uso del internet. También podría deberse a que algunos jefes de hogar no tienen conocimiento sobre el manejo de las TIC. (Ver Tabla N° C9).

- Un jefe de hogar que vivía en una zona urbana tenía un 21.56% más probabilidad de usar el internet, en comparación a un jefe de hogar que vivía en una zona rural. En ese contexto, a nivel de todo el Perú, los jefes de hogar que vivían en la zona urbana tenían mayores posibilidades de hacer el uso de internet, frente a aquellos jefes de hogar que vivían en la zona rural. Esto indica que en la zona urbana la infraestructura de las redes de conexión a internet está distribuida casi en todo su territorio, además los jefes de hogar en la zona urbana tienen mayores conocimientos sobre el manejo de las TIC.

#### **4.1.4. Presentación de los resultados para el objetivo específico 3**

En la sección de materiales y métodos se presentó el modelo econométrico que se utilizó para alcanzar el objetivo específico 3, el cual está conformado de una variable de resultado, una variable de tratamiento y 15 variables de control o

variables explicativas. Para calcular el impacto del uso de internet sobre el empleo, se utilizó el estimador del Efecto Promedio de Tratamiento en los Tratados, más conocido por sus iniciales en inglés como el ATT, tal como se muestra a continuación:

$$ATT = E(Y_{1i}|P(X), U_i = 1) - E(Y_{0i}|P(X), U_i = 0) \quad \dots \text{Ecuación 50}$$

#### 4.1.4.1. Estimación de la puntuación de propensión

Al estimar el puntaje de propensión, se deben tomar dos decisiones. El primero se refiere al modelo a utilizar para la estimación y el segundo a las variables a incluir en este modelo. En ese contexto, se eligió el modelo Logit, conjuntamente con las 15 covariables (como las variables de control) y la variable uso de internet (como la variable de tratamiento), debido a que el mejor modelo según el objetivo específico 2 fue el modelo Logit, el cual se puede visualizar en la Tabla N° 18.

**Tabla 18**

*Coefficientes y medidas de ajuste del modelo Logit del proceso PSM*

<b>Variabes de control y medidas de ajuste</b>	<b>Modelo Logit Uso de internet (Coeficiente)</b>
Edad	-0.0758*** (-44.59)
Genero	0.2179*** (4.1)
Est_civil	-0.2652*** (-4.94)
Primaria	1.5086*** (6.89)
Secundaria	2.6521*** (12.16)
Sup_no_univ	3.9153*** (17.49)
Sup_univ	4.6467***



<b>VARIABLES DE CONTROL Y MEDIDAS DE AJUSTE</b>	<b>Modelo Logit Uso de internet (Coeficiente)</b>
	(20.18)
Maestria	6.8909*** (13.66)
Miembros	-0.1082*** (-8.14)
Estudiantes	0.1132** (2.41)
Celular	2.6501*** (15.87)
Computadora	1.0276*** (23.19)
Sierra	-1.053*** (-24.06)
Selva	-0.8885*** (-18.52)
Urbano	0.9194*** (22.45)
Modelo completo con Log-Lik	-10169.257
LR $\chi^2(15)$	16232.037
Prob > $\chi^2$	0.000
$R^2$ de McFadden	0.444
$R^2$ de recuento	0.820
Área bajo la curva ROC	0.907

-Los números entre paréntesis son los estadísticos z.

-Los asteriscos indican el nivel de significancia: \*\*\* al 1% y \*\* al 5%.

Fuente: Elaboración propia

Como era de esperarse en la Tabla N° 18 se observa que todas las variables de control son significativas al 1%, a excepción de la variable estudiantes que es significativo al 5%. Además, se evidencia que existe una significancia conjunta en el modelo Logit, debido a que el estadístico LR es mayor al 5%. Por otra parte, se muestra que el modelo tiene un buen ajuste, ya que la cantidad de predicciones correctas que realizó el modelo sobre las puntuaciones de propensión es del 82% y el área bajo la curva ROC es del 90.7%.



Una vez analizado el modelo a groso modo y pasar las pruebas estadísticas, es posible utilizar los propensity score o puntuaciones de propensión para realizar el emparejamiento mediante las distintas métricas que existen en la literatura.

#### **4.1.4.2. Elección de un algoritmo coincidente**

Todos los estimadores coincidentes contrastan el resultado de un individuo tratado con los resultados de los individuos del grupo de comparación. Por lo tanto, para la presente investigación se utilizó 3 algoritmos de coincidencia, las cuales vienen establecidas en el paquete del psmatch2 en STATA. Los algoritmos mencionados son:

- Métrica de emparejamiento o coincidencia del vecino más cercano.
- Métrica de emparejamiento o coincidencia de radio y calibre.
- Métrica de emparejamiento o coincidencia de kernel.

#### **4.1.4.3. Verificación del soporte común**

El estimador del ATT se define en la región de soporte común, por lo que un paso importante es determinar la superposición y la región del soporte común entre el grupo de tratamiento y comparación. En ese sentido, se ha establecido el soporte común en base a las puntuaciones de propensión estimadas por el modelo Logit, tal como se muestra en la Tabla N° 19.

**Tabla 19***Definición del soporte común*

<b>Métrica de emparejamiento</b>	<b>Soporte común</b>
Vecino más cercano	0.0000489 - 0.9996326
Radio y calibre	0.0000489 - 0.9996326
Kernel	0.0000489 - 0.9996326

Fuente: Elaboración propia

Como se puede observar en la tabla anterior, el soporte común se ha definido entre las puntuaciones de propensión que se encuentran entre 0.0000489 y 0.9996326.

#### 4.1.4.4. Estimación del impacto

Para estimar el impacto se utilizó el estimador del ATT en base al PSM, que es simplemente la diferencia media en los resultados sobre el apoyo común, ponderada adecuadamente por la distribución del puntaje de propensión de los participantes. Por tanto, en la Tabla N° 20 se presenta los impactos obtenidos por las 3 métricas de emparejamiento. Cabe aclarar que en la métrica del vecino más cercano se consideró 5 vecinos más cercanos, mientras que en la métrica de radio y calibre se estableció una distancia del 10%.

**Tabla 20***Efecto promedio de tratamiento en los tratados (ATT)*

<b>Métrica de emparejamiento</b>	<b>Grupos de tratamiento</b>	<b>Grupo de control</b>	<b>ATT</b>	<b>T estadístico</b>
Vecino más cercano	6.6496	6.2942	0.3554***	11.79
Radio y calibre	6.6496	6.2696	0.3801***	19.68
Kernel	6.6496	6.2874	0.3622***	16.28

-Los asteriscos indican el nivel de significancia: \*\*\* al 1% y \*\* al 5%.

Fuente: Elaboración propia



En la tabla anterior se evidencia que los impactos estimados son significativos, debido a que los valores del t-estadístico del ATT estimado por todas las métricas de emparejamiento son igual a 11.79, 19.68 y 16.28, respectivamente, mayores a 1.96 en valor absoluto, lo cual muestra que estos valores caen en una zona de rechazo de la distribución t de student, por lo tanto, es posible rechazar la hipótesis nula ( $ATT=0$ ), indicando de esta manera que el ATT estimado es estadísticamente diferente de 0.

Para interpretar los impactos estimados, se debe tener en cuenta que el ATT en este caso es la diferencia promedio entre el resultado potencial en relación a la variable ingreso per cápita de los jefes de hogar del Perú que usaron el internet (grupo de tratamiento) y el resultado potencial en relación a la variable ingreso per cápita de los jefes de hogar del Perú que no usaron el internet (grupo de control), en el año 2020. En ese sentido, se procede a la interpretación del ATT para todas las métricas de emparejamiento empleadas en esta investigación:

- Respecto al ATT estimado con la métrica de emparejamiento del vecino más cercano, se obtuvo que, en promedio, los jefes de hogar que usaron el internet ganaron mensualmente un 35.54% más que los jefes de hogar que no usaron el internet, lo cual indicaría que, el hecho de que los jefes de hogar hicieron el uso de internet aumentó su ingreso per cápita mensual en un 35.54%, en comparación de los jefes de hogar que no usaron el internet.
- Referente al ATT estimado con la métrica de emparejamiento de radio y calibre, se obtuvo que, en promedio, los jefes de hogar que



usaron el internet ganaron mensualmente un 38.01% más que los jefes de hogar que no usaron el internet, lo cual indicaría que, el hecho de que los jefes de hogar hicieron el uso de internet aumentó su ingreso per cápita mensual en un 38.01%, en comparación de los jefes de hogar que no usaron el internet.

- En relación al ATT estimado con la métrica de emparejamiento de kernel, se obtuvo que, en promedio, los jefes de hogar que usaron el internet ganaron mensualmente un 36.22% más que los jefes de hogar que no usaron el internet, lo cual indicaría que, el hecho de que los jefes de hogar hicieron el uso de internet aumentó su ingreso per cápita mensual en un 36.22%, en comparación de los jefes de hogar que no usaron el internet.

Por lo tanto, de acuerdo a los resultados obtenidos a través del estimador del ATT, en base a las métricas de emparejamiento mencionados anteriormente, se evidencia que el uso de internet genera un impacto positivo en el ingreso y por ende en el empleo, ya que el ingreso es un indicador o una variable proxi del empleo, y por la misma razón de que un individuo que está percibiendo ingresos, significa que esta empleado o se encuentra trabajando, ya sea formalmente y/o informalmente. Entonces, con estos resultados se puede afirmar que, en el año 2020, los jefes de hogar que sabían usar el internet a nivel de todo el país, en promedio, percibieron mayores ingresos, respecto a los jefes de hogar que no sabían usar el internet.

#### 4.1.4.5. Prueba de la propiedad de balanceo

Dado que no se condicionó a todas las variables explicativas sino al puntaje de propensión, se debe verificar si el procedimiento de emparejamiento es capaz de equilibrar la distribución de las variables relevantes tanto en el grupo de control como en el grupo de tratamiento. La idea básica es comparar la situación antes y después del emparejamiento y verificar si persisten diferencias después de condicionar la puntuación de propensión, a fin de evitar la subestimación o sobreestimación de los impactos y cumplir con el supuesto de independencia condicional del PSM. En ese contexto, la Tabla N° 21 muestra la diferencia de medias o el sesgo de las variables explicativas (bias) del grupo de tratamiento y control, después de emparejar.

**Tabla 21**

*Test de equilibrio*

Variables de control	Vecino más cercano			Radio y calibre			Kernel		
	%bias	t	p> t	%bias	t	p> t	%bias	t	p> t
Edad	10.2	8.67	0.000	-3.7	-3.19	0.001	0.8	0.71	0.478
Genero	-7.7	-6.09	0.000	-3.4	-2.64	0.008	-4.5	-3.53	0.000
Est_civil	-1.2	-0.94	0.347	-2.3	-1.82	0.068	-2.1	-1.64	0.101
Primaria	-0.1	-0.09	0.932	-1.7	-1.76	0.078	-0.6	-0.67	0.506
Secundaria	-1.5	-1.11	0.267	-11.6	-8.75	0.000	-2.6	-2.00	0.046
Sup_no_univ	-5.8	-3.47	0.001	-10.5	-6.16	0.000	-15.9	-9.24	0.000
Sup_univ	1.1	0.66	0.512	20.8	12.97	0.000	11.5	6.92	0.000
Maestria	15	9.92	0.000	24.8	18.69	0.000	21.6	15.58	0.000
Miembros	-4.9	-4.27	0.000	-6.2	-5.40	0.000	-5.9	-5.22	0.000
Estudiantes	-7.3	-5.98	0.000	-4.1	-3.31	0.001	-4.7	-3.84	0.000
Celular	-0.4	-1.07	0.286	1.1	2.75	0.006	0.1	0.23	0.815
Computadora	9.2	6.02	0.000	16.2	10.67	0.000	11.1	7.31	0.000
Sierra	4.8	4.03	0.000	-0.8	-0.68	0.496	2.5	2.11	0.035
Selva	0.9	0.72	0.475	1.5	1.26	0.206	3	2.58	0.010
Urbano	3.1	2.61	0.000	4.5	3.79	0.000	2.6	2.21	0.027

Fuente: Elaboración propia



De acuerdo a los resultados mostrados en la tabla anterior, se verifica que no se cumple con la propiedad de equilibrio en la mayoría de las variables explicativas, debido a que los valores del t-estadístico de estas variables explicativas son mayores a 1.96 en valor absoluto, lo cual indica que estos valores caen en una zona de rechazo de la distribución t de student, por lo tanto, se debe rechazar la hipótesis nula (diferencia de medias del grupo de tratamiento – diferencia de medias del grupo de control = 0), comprobando de esta manera que las medias de la mayoría de las variables explicativas entre ambos grupos son estadísticamente diferentes de cero, es decir, las medias no son estadísticamente iguales. Por otra parte, las probabilidades de estas medias en muchas variables explicativas son menores al 5%, lo que también muestra que no se cumple con la propiedad de equilibrio. Por ejemplo, en la métrica de emparejamiento de radio y calibre solo las variables estado civil, primaria, sierra y selva estarían cumpliendo con la propiedad de equilibrio, las demás no cumplirían, análogamente en las demás métricas de emparejamiento sucede algo similar. Ante esta situación, una posible solución según la literatura sería quedarse solo con las variables que cumplen la propiedad de equilibrio, sin embargo, esto conllevaría a que los resultados del ATT sean sesgados, ya que el simple hecho de que existan pocas variables que expliquen el uso de internet se perdería mucha información y las puntuaciones de propensión no serían muy útiles para emparejar. Por lo tanto, probar otra especificación del modelo no sería la mejor alternativa, a lo contrario ocasionaría más sesgo de la que podría tener.



En este contexto, se puede afirmar que realmente si existe un impacto del uso de internet en el ingreso y por ende en el empleo, sin embargo, estos resultados podrían estar sesgado, es decir, podría existir una subestimación o sobreestimación del valor estimado del impacto, ya que no se cumple con la propiedad de equilibrio en muchas variables explicativas y por ende no se estaría cumpliendo con el supuesto de independencia condicional del PSM.

## 4.2. DISCUSIÓN

Una vez obtenido los resultados para cada objetivo específico, es indispensable comparar estos resultados con la teoría y con los resultados que obtuvieron otros autores en sus investigaciones. En ese sentido, se procede a contrastar los resultados que se obtuvieron en esta investigación con la literatura revisada.

Respecto al objetivo específico 1, se pudo identificar que las covariables como: material, sierra y selva son los factores que influyen negativamente en la probabilidad de que un hogar tenga acceso a internet disponible, mientras que las covariables como: habitaciones, electricidad, estudiantes, celular, computadora y urbano son los factores que influyen positivamente en la probabilidad de que un hogar tenga acceso a internet disponible, contrastando de esta manera con la hipótesis específica 1, a excepción de las covariables: miembros y niños, que no salieron significativas al momento de estimar sus parámetros. Por lo tanto, se pudo evidenciar que los hogares con mayor número de habitaciones tuvieron mayores probabilidades de tener conexión a internet disponible, es decir, por cada habitación adicional que tenía un hogar, su probabilidad de tener conexión a internet disponible se incrementaba en un 3.13%. Por otro lado, los hogares que fueron contruidos con material precario, los hogares que se ubicaban en la sierra y selva, tenían



un 13.81%. 11.88% y 5.51% menos probabilidades de tener conexión a internet disponible, respectivamente, en comparaciones a sus categorías base. Además, los hogares que tenían conexión a energía eléctrica, los hogares en donde existían estudiantes, los hogares en donde algunos de sus miembros disponían de celular, los hogares en donde algunos de sus miembros disponían de computadora y/o laptop, y los hogares que se ubicaban en la zona urbana tuvieron un 8.42%, 4.77%, 26.62%, 44.44% y 17.64% más de probabilidades de tener conexión a internet disponible o fijo.

Comparando con la teoría de los recursos y de la apropiación de nuevas tecnologías desarrollada por Van Dijk (2012), el acceso físico y material a internet está determinada por la motivación, y este a la vez está determinada por los factores económicos, sociales y culturales, tales como: el factor temporal (tener tiempo para utilizar medios digitales), el factor material (posesión e ingresos), el factor mental (capacidad técnica, motivación), el factor social (tener una red social para ayudar en el uso de los medios digitales) y el factor cultural (status y gusto por estar en el mundo de los medios digitales). En este contexto, en la presente investigación se pudo comprobar de alguna u otra manera esta teoría, ya que para que los miembros de un hogar se motiven para hacerse conectar el internet en casa desde una red pública, primero deben estar motivado en adquirir los medios digitales (celular, computadora y/o laptop), es decir, primero está la motivación por la adquisición de medios digitales y segundo esta la motivación por la decisión de contar con internet en casa.

De igual forma, en el estudio que realizó Grazzi (2011), también encontró que los hogares con mayor densidad de estudiantes tenían mayores probabilidades de tener conexión a internet en casa, mientras que los hogares que se encontraban en el área rural tenían menores probabilidades de tener conexión a internet en casa. De igual modo, en la



investigación realizada por Humpiri (2022), obtuvo que el número de miembros en el hogar no fue significativo, además, en el estudio realizado por Fernández & Medina (2011), también obtuvieron que la presencia de niños menores a los 10 años en el hogar, no fueron significativos, muy similar al resultado que se obtuvo en la presente investigación.

En relación al objetivo específico 2, se pudo identificar que las covariables como: edad, estado civil, miembros, sierra y selva son los factores que influyen negativamente en la probabilidad de que un jefe de hogar use el internet, mientras que las covariables como: genero, primaria, secundaria, superior no universitaria, superior universitaria, maestría y doctorado, estudiantes, celular, computadora y/o laptop y urbano son los factores que influyen positivamente en la probabilidad de que un jefe de hogar use el internet, contrastando de esta manera con la hipótesis específica 2, a excepción de la covariable niños, que no salió significativa al momento de estimar su parámetro. Por lo tanto, se pudo evidenciar que los jefes de hogar que tenían mayor edad y mayor número de miembros tuvieron menores probabilidades de usar el internet, es decir, por cada año adicional y por cada miembro adicional que tenía un jefe de hogar, su probabilidad de usar el internet disminuía en 1.83% y 2.60%, respectivamente. Por otro lado, los jefes de hogar que se encontraban en condición de casado o conviviente, los jefes de hogar que vivían en la sierra y los jefes de hogar que vivían en la selva, tenían un 6.32%, 25.29% y 21.58% menos probabilidades de usar el internet, sucesivamente, en comparación a sus categorías base. Además, los jefes de hogar que eran hombres, que eran estudiantes y/o tenían miembros estudiantes, que tenían celular, que tenían computadora y que vivían en la zona urbana, tenían un 5.03%, 2.68%, 39.45%, 24.87% y 21.56% más probabilidades de usar el internet, respectivamente, en comparación a sus categorías base, asimismo, los jefes de hogar que habían estudiado primaria, secundaria, superior no universitaria,



superior universitaria y maestría y doctorado, tenían un 14.31%, 37.90%, 67.72%, 79.83% y 93.18% mayores probabilidades de usar el internet, respectivamente, en comparación a sus categorías base.

Comparando con la teoría de los recursos y de la apropiación de nuevas tecnologías de Van Dijk (2012), el uso de las TIC está determinada por 3 etapas previas de la brecha digital (motivación, acceso físico y material y habilidades digitales), en ese orden, uno antecediendo al otro. Y cada una de estas etapas están determinada a la vez por las desigualdades categóricas personales (edad, genero, raza, inteligencia, personalidad y salud) y desigualdades categóricas posicionales (posición laboral, educación, tamaño de hogar y nación), asimismo, está determinada por los factores económicos, sociales y culturales, tales como: el factor temporal (tener tiempo para utilizar medios digitales), el factor material (posesión e ingresos), el factor mental (capacidad técnica, motivación), el factor social (tener una red social para ayudar en el uso de los medios digitales) y el factor cultural (status y gusto por estar en el mundo de los medios digitales). En este contexto, en el presente estudio de alguna u otra manera se estaría comprobando esta teoría, debido a que la brecha digital para los jefes de hogar en el Perú se ha relacionado con la edad, genero, educación, estado civil, etc., además, se ha relacionado con la motivación, el cual es una etapa muy importante para que un jefe de hogar adquiera los medios digitales (celular, computadora y/o laptop) y con ello haga el uso de internet.

De la misma manera, comparando con el estudio que realizaron (Al-Hammadany & Heshmati, 2011) se identificó que las personas jóvenes tienen mayores probabilidades de usar el internet, es decir, las personas que se encontraron en la categoría de 25 a 40 años y en la categoría de 40 años a más, tuvieron un 1.1% y 13.5% menos probabilidades



de usar el internet que las personas que se encontraron en la categoría de 12 a 25 años, asimismo, Humpiri (2022), mostró que si la edad del jefe del hogar aumenta en 1 año adicional, su probabilidad de que use el internet aumenta en 1.25%. Por otro lado, en el estudio de Botello (2014), se encontró que los hombres, las personas que habitan en la región central y las personas que residen en la zona urbana tenían un 4.2%, 20% y 10% más probabilidades de usar el internet, consecutivamente, respecto de sus categorías base, mientras que referente a la educación, las personas que tenían títulos de posgrado, educación universitaria, educación secundaria y educación primaria tuvieron un 84%, 83%, 38% y 13% más probabilidades de usar internet, respectivamente, que una persona que no tenga ningún grado educativo. Además, en la investigación de Siaw et al. (2020), se evidenció que los que tienen una educación secundaria o superior tienen un 6.14% más probabilidades de usar el internet en comparación de los que tienen educación primaria o ningún nivel educativo y los que perciben que navegar con un teléfono y/o computadora es fácil tienen un 2.23% más probabilidades de utilizar el internet. Del mismo modo, en la investigación de Noce & McKeown (2007), se encontró que el número de niños en el hogar tampoco fue significativo, al igual que en el presente estudio.

En cuanto, al objetivo específico 3, a través de las métricas de emparejamiento del vecino más cercano, radio y calibre, y kernel se obtuvo que los jefes de hogar que usaron el internet percibieron, en promedio, un 35.54%, 38.01% y 36.22% más ingreso per capital mensual, respectivamente, en comparación a los jefes de hogar que no usaron el internet, mostrando de esta manera que el uso de internet genera un impacto positivo sobre el ingreso y por ende sobre el empleo, el cual resultó ser, tal como se planteó en la hipótesis específica 3 en un comienzo. De la misma forma, en el estudio que realizó Fernández & Medina (2011), se encontró que el uso de las TIC genera un impacto positivo sobre los ingresos, evidenciando que el acceso conjunto a estos medios puede incrementar



el ingreso promedio per cápita del hogar en 105 nuevos soles. Además, en el estudio de Navarro (2010), también se encontró un impacto positivo del uso de internet sobre los ingresos de los trabajadores asalariados y los trabajadores por cuenta propia, el cual fue del 18% y el 30%, respectivamente. Asimismo, en el estudio que realizó De Los Ríos (2010), se encontró que el impacto de la adopción de internet en el cambio en los ingresos laborales fue de 428 nuevos soles, equivalentes a 6% del ingreso promedio de los trabajadores, mostrando de esta manera un impacto positivo.

Del mismo modo, la postura de Mantecón (2023), indica que las TIC crean nuevos empleos y mejoran la eficiencia y la productividad de las industrias, en relación al trabajo autoprogramable, sin embargo, para este tipo de trabajo se necesitan personas capaces de adaptarse a los distintos cambios de la tecnología y que tengan un alto grado de conocimiento. A pesar de eso, el uso de las TIC genera nuevos empleos en el Perú y en el mundo entero.

Cabe aclarar que, la teoría neoclásica plantea que la intervención del estado y de los sindicatos provocan el aumento de los salarios reales por encima del nivel de equilibrio y por tanto, el desempleo (Colás, 2007). Por lo tanto, dado que esta teoría evidencia una relación entre el empleo y salario, entonces se optó por utilizar la variable ingresos como una variable proxy del empleo.

En síntesis, existe una similitud entre los resultados obtenidos en esta investigación y los que se obtuvieron en otras investigaciones, ya que en este estudio se evidencia un impacto positivo del uso de internet sobre el empleo, evidenciado a través del ingreso per cápita de los jefes de hogar, el cual se consideró como una variable proxy del empleo. De la misma forma, en otros estudios se evidencian un impacto positivo del



uso de internet sobre los ingresos, pero con diferentes magnitudes, debido a que los estudios se realizaron bajo otros contextos y en diferentes años.



## V. CONCLUSIONES

- En la presente investigación se ha identificado los factores que influyen en la probabilidad de que un hogar tenga acceso a internet disponible, donde se ha utilizado el modelo Probit por tener el mejor ajuste y para estimar los parámetros de las covariables se ha utilizado el método de la máxima verosimilitud. Por otra parte, se ha determinado los factores que influyen en la probabilidad de que un jefe de hogar use internet, para lo cual se ha utilizado el modelo Logit por tener el mejor ajuste y para estimar los parámetros de las covariables se ha utilizado de igual manera el método de la máxima verosimilitud. Además, se ha estimado el impacto que genera el uso de internet sobre el empleo, para lo cual se ha utilizado el método de evaluación de impacto denominado Propensity Score Matching (PSM), conjuntamente con sus distintas métricas de emparejamiento, tales como: el vecino más cercano, radio y calibre, y kernel, los cuales sirvieron para estimar el efecto promedio de tratamiento en los tratados (ATT).
- Los factores que influyen en la probabilidad de que un hogar tenga acceso a internet disponible fueron: el material de vivienda, el número de habitaciones, la energía eléctrica, la existencia de estudiantes, la existencia de miembros con celular, la existencia de miembros con computadora y/o laptop, la región geográfica y la zona de ubicación. En consecuencia, en el año 2020, a nivel de todo el Perú, se pudo evidenciar que los hogares con mayor número de habitaciones tuvieron mayores probabilidades de tener conexión a internet disponible, es decir, por cada habitación adicional que tenía un hogar, su probabilidad de tener conexión a internet disponible se incrementaba en un 3.13%. Por otro lado, los hogares que fueron construidos con material precario, los hogares que se ubicaban en la sierra y los hogares que se



ubicaban en la selva, tenían un 13.81%, 11.88% y 5.51% menos probabilidades de tener conexión a internet disponible, respectivamente, en comparación a sus categorías base. Además, los hogares que tenían conexión a energía eléctrica, los hogares en donde existían estudiantes, los hogares en donde alguno de sus miembros disponía de celular, los hogares en donde alguno de sus miembros disponía de computadora y/o computadora, y los hogares que se ubicaban en la zona urbana, tuvieron un 8.42%, 4.77%, 26.62%, 44.44% y 17.64% más probabilidades de tener conexión a internet disponible, respectivamente, en comparación a sus categorías base.

- Los factores que influyen en la probabilidad de que un jefe de hogar use internet fueron: la edad, el género, el estado civil, la educación, el tamaño de familia, la existencia de estudiantes, la disponibilidad de celular, la disponibilidad de computadora y/o laptop, la región geográfica y la zona de residencia. Por consiguiente, en el año 2020, a nivel de todo el Perú, se pudo evidenciar que los jefes de hogar que tenían mayor edad y mayor número de miembros tuvieron menores probabilidades de usar el internet, es decir, por cada año adicional y por cada miembro adicional que tenía un jefe de hogar, su probabilidad de usar el internet disminuía en 1.83% y 2.60%, respectivamente. Por otro lado, los jefes de hogar que se encontraban en condición de casado o conviviente, los jefes de hogar que vivían en la sierra y los jefes de hogar que vivían en la selva, tenían un 6.32%, 25.29% y 21.58% menos probabilidades de usar el internet, sucesivamente, en comparación a sus categorías base. Además, los jefes de hogar que eran hombres, que eran estudiantes y/o tenían miembros estudiantes, que tenían celular, que tenían computadora y que vivían en la zona urbana, tenían un 5.03%, 2.68%, 39.45%, 24.87% y 21.56% más probabilidades de usar el internet, respectivamente, en



comparación a sus categorías base, asimismo, los jefes de hogar que habían estudiado primaria, secundaria, superior no universitaria, superior universitaria y maestría y doctorado, tenían un 14.31%, 37.90%, 67.72%, 79.83% y 93.18% mayores probabilidades de usar el internet, respectivamente, en comparación a los jefes de hogar que solo habían estudiado el nivel inicial o no habían estudiado algún nivel educativo.

- El uso de internet generó un impacto positivo sobre el ingreso y por ende sobre el empleo, debido a que a través de las métricas de emparejamiento del vecino más cercano, radio y calibre, y kernel se obtuvo que, en el año 2020, a nivel de nacional, los jefes de hogar que usaron el internet, en promedio, percibieron un 35.54%, 38.01% y 36.22% más ingreso per capital mensual, respectivamente, en comparación a los jefes de hogar que no usaron el internet, evidenciando de esta manera que los jefes de hogar que sabían usar el internet, en promedio, percibieron mayores ingresos, respecto a los jefes de hogar que no sabían usar el internet.



## VI. RECOMENDACIONES

De acuerdo a las conclusiones que se tuvo para cada objetivo específico, en base a los resultados obtenidos en la presente investigación, se recomienda los siguientes aspectos muy importantes:

- Al Gobierno Nacional, a través del Ministerio de Desarrollo e Inclusión Social (MIDIS), se recomienda implementar políticas de subsidio monetario en el pago de los servicios de internet, a fin de incrementar las posibilidades de acceso a internet fijo y/o móvil en los hogares del Perú. Asimismo, a las autoridades competentes del Ministerio de Transportes y Comunicaciones (MTC), a través del Fondo de Inversión en Telecomunicaciones (FITEL), se sugiere implementar más proyectos o programas de inversión relacionadas a la ampliación de infraestructura de la banda ancha de internet, sobre todo en los lugares más alejados del Perú, en el marco a la Ley N°29904, Ley de promoción de la banda ancha y construcción de la red dorsal nacional de fibra óptica.
- Además, a los distintos niveles de gobierno (nacional, regional y local) se recomienda desarrollar estrategias y plantear políticas para reducir la brecha digital referente al uso de internet mediante las capacitaciones y asistencia técnica, fortaleciendo de esta manera las habilidades y el conocimiento de las personas del Perú sobre el manejo de internet. Asimismo, al Organismo Supervisor de Inversión Privada en Telecomunicaciones (OSIPTEL), se recomienda regular y supervisar la calidad de los servicios que brinda las empresas telefónicas referente al internet, a fin de que los usuarios disfruten de este servicio de manera amena y satisfactoria.



- Por otra parte, se recomienda al Ministerio de Trabajo y Promoción del Empleo (MTPE) para que implemente un programa de alfabetización digital a nivel nacional, a fin de incrementar la magnitud del impacto del uso de internet sobre el empleo, asimismo, se sugiere al Ministerio de Educación (MINEDU) para que impulse nuevos cursos en la malla curricular de las instituciones educativas públicas y privadas (inicial, primaria, secundaria y superior) sobre el manejo e importancia de las TIC, a fin de que los estudiantes de diferentes niveles en un futuro tengan mayores oportunidades para conseguir un empleo. Además, a los diferentes niveles de gobierno (nacional, regional y local) se recomienda que realicen capacitaciones o talleres de concientización para que las personas del Perú den el buen uso al internet, tales como: buscar empleo, trabajar por online, crear páginas web para sus emprendimientos, buscar información importante, etc., asimismo, se sugiere que impulsen el teletrabajo, ya que en este estudio se evidenció la gran importancia que tiene el uso de internet en el empleo.
- Por último, se recomienda a las futuras investigaciones a que realicen un estudio de evaluación de impacto considerando 2 periodos, antes y después de la pandemia del COVID-19, a fin de estimar el impacto real del uso de internet sobre el empleo en tiempos de pandemia, asimismo, se sugiere a que realicen un estudio exclusivamente sobre el uso de internet y su impacto en el empleo, asignando como grupo de tratamiento a los usuarios que buscan trabajo por internet y como grupo de control a los usuarios que buscan trabajo por otros medios, y evaluando el impacto sobre los usuarios que realmente han encontrado un trabajo por estos medios, ya que los resultados que se obtendrían serían más robustos y reales.



## VII. REFERENCIAS BIBLIOGRÁFICAS

- Al-Hammadany, F. H., & Heshmati, A. (2011). Determinants of internet use in Iraq. *International Journal of Communication*, 5(2011), 1967-1989.  
<https://ijoc.org/index.php/ijoc/article/view/746/670>
- Amaya, J. (2010). *Sistemas de información gerenciales: Hardware, software, redes, internet, diseño* (2.<sup>a</sup> ed.). Ecoe Ediciones.
- Arias, W. (2023). *El uso del internet y el aprendizaje en los estudiantes de la Escuela Profesional de Educación Primaria de la UNA, 2022* [Tesis de pregrado]. Universidad Nacional del Altiplano.  
<https://repositorio.unap.edu.pe/handle/20.500.14082/20134>
- Asociación Latinoamericana de Integración. (2003). *La brecha digital y sus repercusiones en los países miembros de la ALADI*.  
<https://repositorio.aladi.org/handle/20.500.12909/16801>
- Augurzky, B., & Schmidt, C. (2001). The propensity score: A means to an end. *IZA Discussion Papers*, (271). <https://hdl.handle.net/10419/21122>
- Banco Central de Reserva del Perú. (2011). *Glosario de términos económicos*.  
<https://www.bcrp.gob.pe/docs/Publicaciones/Glosario/Glosario-BCRP.pdf>
- Barrios, J. F. (2021). *Determinantes de acceso y de uso de internet de los hogares en cabeceras y centros poblados en Colombia a partir de la Encuesta de Calidad de Vida de 2019* [Tesis de maestría]. Universidad Nacional de Colombia.
- Black, D., & Smith, J. (2004). How robust is the evidence on the effects of the college quality? Evidence from Matching. *Journal of Econometrics*, 121(1-2), 99-124.  
<https://doi.org/10.1016/j.jeconom.2003.10.006>
- Botello, H. A. (2014). Determinantes del acceso a internet en Colombia. *Ánfora*, 21(37), 21-36. <https://www.redalyc.org/pdf/3578/357833888001.pdf>
- Bryson, A., Dorsett, R., & Purdon, S. (2002). *The use of Propensity Score Matching in the evaluation of labour market policies* (4). <http://eprints.lse.ac.uk/4993/>



- Caliendo, M., & Kopeinig, S. (2005). Some practical guidance for the implementation of Propensity Score Matching. *IZA Discussion Paper*, (1588).  
<https://docs.iza.org/dp1588.pdf>
- Chaudhuri, A., Flamm, K. S., & Horrigan, J. (2005). An analysis of the determinants of internet access. *Telecommunications Policy*, 29(9-10), 731-755.  
<https://doi.org/10.1016/j.telpol.2005.07.001>
- Cochrane, W., & Chambers S. (1965). The planning of observational studies of human populations. *Journal of the Royal Statistical Society*, 128(2), 234-266.  
<https://doi.org/10.2307/2344179>
- Colás, S. (2007). Diferentes enfoques del análisis del empleo en la teoría económica. *Contribuciones a la Economía*. <https://www.eumed.net/ce/2007b/scg.htm>
- Comisión Económica para América Latina y el Caribe. (2013). *Economía digital para el cambio estructural y la igualdad*. <https://hdl.handle.net/11362/35408>
- Comisión Económica para América Latina y el Caribe. (2021). *Datos y hechos sobre la transformación digital*. <https://hdl.handle.net/11362/46766>
- Davis, F. (1989). Perceived usefulness, perceived ease of use and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340.  
<https://doi.org/10.2307/249008>
- Davis, F., Bagozzi, R., & Warshaw, P. (1989). User acceptance of computer technology: a comparison of two theoretical models. *Management sciences*, 35(8), 983-1003. <https://www.jstor.org/stable/2632151>
- De Los Ríos, C. A. (2010). *Impacto del uso de internet en el bienestar de los hogares peruanos: Evidencia de un panel de hogares 2007-2009*.
- Fernández, R., & Medina, P. (2011). *Evaluación del impacto del acceso a las TIC sobre el ingreso de los hogares: Una aproximación a partir de la metodología del Propensity Score Matching y datos de panel para el caso peruano*.  
<https://repositorio.iep.org.pe/handle/IEP/1279>



- Flores, J. J., Hernández, R. M., & Garay, R. (2020). Tecnologías de información: Acceso a internet y brecha digital en Perú. *Revista Venezolana de Gerencia*, 25(90), 504-519. <https://www.redalyc.org/articulo.oa?id=29063559007>
- Fora, A. (2023). *Las TICs y el aprendizaje del idioma inglés en los estudiantes del IX ciclo de Ingeniería Económica de la UNA Puno 2022* [Tesis de segunda especialidad]. Universidad Nacional del Altiplano. <https://repositorio.unap.edu.pe/handle/20.500.14082/19727>
- Gómez, D. A., Alvarado, R. A., Martínez, M., & Díaz de León, C. (2018). La brecha digital: una revisión conceptual y aportaciones metodológicas para su estudio en México. *Revistas UNAM*, 6(16), 47-62. <https://doi.org/10.22201/enesl.20078064e.2018.16.62611>
- Grazzi, M. (2011). Patterns of internet use. *ECLAC*, 41-68. <https://hdl.handle.net/11362/35292>
- Haight, M., Quan-Haase, A., & Corbett, B. (2014). Revisiting the digital divide in Canada: the impact of demographic factors on access to the internet, level of online activity, and social networking site usage. *Information, Communication & Society*, 17(4), 503-519. <https://doi.org/10.1080/1369118X.2014.891633>
- Heckman, J., Ichimura, H., Smith, J., & Todd, P. (1998). Characterizing selection bias using experimental data. *Econometrica*, 66(5), 1017-1098. <https://doi.org/10.2307/2999630>
- Heckman, J., Ichimura, H., & Todd, P. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies*, 64(4), 605-654. <https://doi.org/10.2307/2971733>
- Heckman, J., Lalonde, R., & Smith, J. (1999). The economics and econometrics of active labor market programs. *Handbook of Labor Economics*, 3, 1865-2097. [https://doi.org/10.1016/S1573-4463\(99\)03012-6](https://doi.org/10.1016/S1573-4463(99)03012-6)
- Heinrich, C., Maffioli, A., & Vázquez, G. (2010). A primer for applying Propensity-Score Matching. <http://dx.doi.org/10.18235/0008567>



- Hernández, R., Fernández, C., & Baptista, M. del P. (2014). *Metodología de la investigación* (6.<sup>a</sup> ed.). McGraw Hill / Interamericana Editores, S.A. de C.V.
- Huarag, E. M. (2020). Efectos de la pandemia de COVID-19 sobre la adopción de las TIC en el Perú. *Ius Inkarrí*, 9(9), 491-523.  
<https://revistas.urp.edu.pe/index.php/Inkarri/article/view/3697>
- Humpiri, H. E. (2022). *Acceso a internet y los determinantes socioeconómicos en el Perú, periodo 2016-2019* [Tesis de pregrado]. Universidad Nacional del Altiplano. <https://repositorio.unap.edu.pe/handle/20.500.14082/19368>
- Instituto Nacional de Estadística e Informática. (2019). *Perú: Participación de la población en la actividad económica, 2017*.
- Instituto Nacional de Estadística e Informática. (2020). *Ficha Técnica: Encuesta Nacional de Hogares sobre condiciones de vida y pobreza - 2020*.
- Lavado, P. (2021). *Impacto del COVID-19 en el empleo en el Perú: una aproximación por tipo de contrato y tamaño de empresa*.
- Ley N° 30036: Ley que regula el Teletrabajo. Art. 2. (2006). Congreso de la República. Diario Oficial El Peruano.  
<https://cdn.www.gob.pe/uploads/document/file/356823/30036.pdf?v=15670898>  
83
- Mantecón, M. del C. (2023). *El impacto de las nuevas tecnologías en el empleo*.  
<https://repositorio.comillas.edu/jspui/bitstream/11531/68755/2/TFG-%20Mantecon%20Miranda%2C%20Maria%20del%20Carmen.pdf>
- Martínez, R., & Espinal, K. (2023). *Cómo terminar una tesis* (1.<sup>a</sup> ed.). Biblioteca Nacional del Perú.
- Ministerio de Trabajo y Promoción del Empleo. (2022). *Informe anual del empleo en el Perú 2021*.
- Ministerio de Transportes y Comunicaciones. (2020). *Impacto económico del acceso a internet en los hogares peruanos*.



- Mwirigi, N. D. (2013). *Factors influencing the uptake of internet connectivity in Meru municipality, Meru county in Kenya*. *Asian Journal of Social Sciences and Management Studies*, 7(1), 39–52.  
<https://doi.org/10.20448/journal.500.2020.71.39.52>
- Navarro, L. (2010). *The impact of internet use on individual earnings in Latin America* (11). <https://hdl.handle.net/10419/45678>
- Noce, A. A., & McKeown, L. (2007). A new benchmark for Internet use: A logistic modeling of factors influencing internet use in Canada, 2005. *Government Information Quarterly*, 25(3), 462-476.  
<https://doi.org/10.1016/j.giq.2007.04.006>
- Ordaz, J. A., Melgar, M. del C., & Rubio, C. M. (2011). *Métodos estadísticos y econométricos en la empresa y para finanzas*.
- Organismo Supervisor de Inversión Privada en Telecomunicaciones. (2022). *Los servicios públicos de telecomunicaciones en los hogares peruanos: Encuesta Residencial de Servicios de Telecomunicaciones*.  
<https://repositorio.osiptel.gob.pe/handle/20.500.12630/860>
- Organización Internacional del Trabajo. (2021a). *Mercado laboral peruano: impacto de la COVID-19 y recomendaciones de política*.
- Organización Internacional del Trabajo. (2021b). *Perspectivas sociales y del empleo en el mundo: Tendencias 2021* (1.ª ed.).
- Osabohien, R., Worgwu, H., Rafi, S. K., Adediran, O., Matthew, O., & Aderounmu, B. (2022). Impact of business innovation on future employment in Nigeria. *Wiley*, 1-12. <https://doi.org/10.1002/mde.3629>
- Osio, L. (2010). El Teletrabajo: Una opción en la era digital. *Observatorio Laboral Revista Venezolana*, 3(5), 93-109.  
<https://www.redalyc.org/articulo.oa?id=219014912006>
- Pérez, R. (2018). La sociedad del conocimiento y la sociedad de la información como la piedra angular en la innovación tecnológica educativa. *Revista Iberoamericana*



- para la Investigación y Desarrollo Educativo*, 8(16).  
<https://www.redalyc.org/articulo.oa?id=498159332036>
- Priyatna, M. F. (2022). Does internet usage lead to an increase in household incomes? Indonesian rural case study. *Jurnal Ekonomi Indonesia*, 11(1), 13-24.  
<https://jurnal.isei.or.id/index.php/isei/article/view/69>
- Reyes, O., & Franklin, O. R. (2014). Teoría del bienestar y el óptimo de Pareto como problemas microeconómicos. *Revista Electrónica de Investigación en Ciencias Económicas (REICE)*, 2(3), 217-234. <https://doi.org/10.5377/reice.v2i3.1457>
- Rini, A. N., & Rahadiantino, L. (2020). The role of internet utilization among SMEs on household welfare in Indonesia. *Jurnal Ekonomi Indonesia*, 9(1), 25-37.  
<https://doi.org/10.52813/jei.v9i1.42>
- Rodríguez, F. N. (2007). Generalidades acerca de las técnicas de investigación cuantitativa. *Sistema Institucional de Investigación de Unitec (SIIU)*, 2(1), 9-39.  
<https://dialnet.unirioja.es/servlet/articulo?codigo=4942053>
- Roque, E. C. (2019). *Impacto del acceso y uso del servicio de internet en la zona urbana y rural del Perú, año 2015 – 2017* [Tesis de pregrado]. Universidad Ricardo Palma. <https://repositorio.urp.edu.pe/handle/20.500.14138/2565>
- Rosenbaum, P., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-50.  
<https://doi.org/10.1093/biomet/70.1.41>
- Roy, A. D. (1951). Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers*, 3(2), 135-145. <https://doi.org/10.1093/oxfordjournals.oep.a041827>
- Rubin, D. B. (1974). Estimating causal effects to treatments in randomised and nonrandomised studies. *Journal of Educational Psychology*, 66(5), 688-701.  
<https://doi.org/10.1037/h0037350>
- Ruiz, E. A., & Ortiz, H. (2014). Acceso a internet e impacto en los hogares peruanos. Una evaluación a partir de microdatos. *Revista Redes.com*, (9), 361-375.  
<https://dialnet.unirioja.es/servlet/articulo?codigo=4718659>



- Sianesi, B. (2004). An evaluation of the active labour market programmes in Sweden. *The Review of Economics and Statistics*, 86(1), 133-155.  
<https://www.jstor.org/stable/3211664>
- Siaw, A., Jiang, Y., Ankrah, M., & Agbenyo, W. (2020). The impact of internet use on income: The case of rural Ghana. *Sustainability*, 12(8).  
<https://doi.org/10.3390/su12083255>
- Smith, H. (1997). Matching with multiple controls to estimate treatment effects in observational studies. *Sociological Methodology*, 27(1), 325-353.  
<https://doi.org/10.1111/1467-9531.271030>
- Todd, P. (1999). A practical guide to implementing matching estimators.
- Todd, P. (2008). Matching estimators. *The new palgrave dictionary of economics*.
- Tovar, N. E. (2021). *Impacto del acceso al servicio de internet en los ingresos per cápita de los hogares de la región central, año 2019* [Tesis de pregrado]. Universidad Nacional del Centro del Perú.  
<https://repositorio.uncp.edu.pe/handle/20.500.12894/7146>
- Unión Internacional de Telecomunicaciones. (2022). *Informe de conectividad global 2022*. [https://www.itu.int/dms\\_pub/itu-d/opb/ind/d-ind-global.01-2022-pdf-e.pdf](https://www.itu.int/dms_pub/itu-d/opb/ind/d-ind-global.01-2022-pdf-e.pdf)
- Van Dijk, J. A. G. M. (2012). The evolution of the digital divide: The digital divide turns to inequality of skills and usage. *Digital Enlightenment Yearbook*, 57-75.  
<https://www.utwente.nl/en/bms/vandijk/news/The%20Evolution%20of%20the%20Digital%20Divide/Evolution%20of%20the%20Digital%20Divide%20Digital%20Enlightment%20Yearbook%202012.pdf>
- Van Dijk, J. A. G. M., & Van Deursen, A. J. A. M. (2014). *Digital skills. Unlocking the information society* (M. Thomas, J. P. Gee, & J. Palfrey, Eds.; 1.<sup>a</sup> ed.). Palgrave Macmillan. <https://link.springer.com/book/10.1057/9781137437037>
- Vara, A. A. (2012). *Desde la idea hasta la sustentación: 7 pasos para una tesis exitosa* (3.<sup>a</sup> ed.).



- Worgwu, H., Osabohien, R., Al-Faryan, M. A. S., & Precious, O. I. (2023). ICT- social entrepreneurship nexus and job creation in Nigeria. *International Journal of Management, Economics and Social Sciences*, 12(1), 79-103.  
<https://hdl.handle.net/10419/268830>
- Yong, L. A., Rivas, L. A., & Chaparro, J. (2010). Modelo de aceptación tecnológica (TAM): un estudio de la influencia de la cultura nacional y del perfil del usuario en el uso de las TIC. *Revista de Ciencias Administrativas y Sociales (INNOVAR)*, 20(36), 187-203.  
<https://www.redalyc.org/articulo.oa?id=81819028014>
- Zhang, W., & Liu, X. (2023). The impact of internet on innovation of manufacturing export enterprises: Internal mechanism and micro evidence. *Journal of Innovation & Knowledge*, 8(3), 1-12. <https://www.elsevier.es/en-revista-journal-innovation-knowledge-376-pdf-S2444569X23000732>

## ANEXOS

### ANEXO 1. Matriz de consistencia

PROBLEMAS PROBLEMA GENERAL	OBJETIVOS OBJETIVO GENERAL	HIPOTESIS HIPÓTESIS GENERAL	MÉTODO DE ESTIMACIÓN ECONOMÉTRICA	METODOLOGÍA
¿Cuáles son los factores que inciden en el acceso y uso de internet, y que impacto genera el uso de internet sobre la generación del empleo en los hogares del Perú, en el año 2020?	Analizar los factores que inciden en el acceso y uso de internet, y medir el impacto del uso de internet sobre la generación del empleo en los hogares del Perú, en el año 2020.	Los factores que inciden en el acceso a internet son: el material de vivienda, el número de habitaciones, la energía eléctrica, el número de miembros, la existencia de niños, la existencia de estudiantes, la existencia de miembros con celular, la existencia de miembros con computadora y/o laptop, la región geográfica y la zona de ubicación. Mientras que los factores que inciden en el uso de internet son: la edad, el género, el estado civil, la educación, el tamaño de familia, la existencia de niños, la existencia de estudiantes, la disponibilidad de celular, la disponibilidad de computadora y/o laptop, la región geográfica y la zona de residencia; y consecuentemente el uso de internet genera un impacto positivo el empleo en los hogares del Perú, en el año 2020.	Método de la máxima verosimilitud para los modelos Logit y Probit.  Método de evaluación de impacto del Propensity Score Matching (PSM), con la métrica del vecino más cercano, radio y calibre y kernel.	El enfoque de la investigación es cuantitativo.  El alcance de la investigación es explicativo.  El diseño de la investigación es cuasiexperimental.  El tipo de la investigación es aplicada.  El método de la investigación es deductivo.  La unidad de análisis son los hogares del Perú que existieron en el año 2020.  La población según el INEI consta de 14,742,326 hogares.  La muestra según el INEI consta de 53,423 hogares.  La técnica de recolección de datos según el INEI es la encuesta, pero según la presente investigación es el análisis de datos secundarios.
PROBLEMAS ESPECÍFICOS	OBJETIVOS ESPECÍFICOS	HIPÓTESIS ESPECÍFICAS	MÉTODO DE ESTIMACIÓN ECONOMÉTRICA	
¿Cuáles son los factores que influyen en la probabilidad de que un hogar tenga acceso a internet disponible?	Identificar los factores que influyen en la probabilidad de que un hogar tenga acceso a internet disponible.	Los factores que influyen en la probabilidad de que un hogar tenga acceso a internet disponible son: el material de vivienda, el número de habitaciones, la energía eléctrica, el número de miembros, la existencia de niños, la existencia de estudiantes, la existencia de miembros con celular, la existencia de miembros con computadora y/o	Método de la máxima verosimilitud para los modelos Logit y Probit.	El instrumento de recolección de datos según el INEI es cuestionario, pero según la presente investigación es la ficha bibliográfica.  La fuente de información de la



		laptop, la región geográfica y la zona de ubicación.		investigación es la base de datos del ENAHO.
¿Cuáles son los factores que influyen en la probabilidad de que un jefe de hogar use internet?	Determinar los factores que influyen en la probabilidad de que un jefe de hogar use internet.	Los factores que influyen en la probabilidad de que un jefe de hogar use internet son: la edad, el género, el estado civil, la educación, el tamaño de familia, la existencia de niños, la existencia de estudiantes, la disponibilidad de celular, la disponibilidad de computadora y/o laptop, la región geográfica y la zona de residencia.	Método de la máxima verosimilitud para los modelos Logit y Probit.	El proceso de análisis de datos y la estimación se realizó en el software estadístico Stata/MP 16.0.
¿Qué impacto genera el uso de internet sobre el empleo?	Estimar el impacto que genera el uso de internet sobre el empleo.	El uso de internet genera un impacto positivo sobre el empleo.	Método de evaluación de impacto del Propensity Score Matching (PSM), con la métrica del vecino más cercano, radio y calibre y kernel.	

## ANEXO 2. Matriz de operacionalización de variables para el O.E.1.

VARIABLES	DEFINICIÓN CONCEPTUAL	DEFINICIÓN OPERACIONAL	DIMENSIONES	INDICADORES	ITEMS
<b>Variable 1:</b> <b>Acceso a internet</b> (Variable dependiente)	Es la disposición o decisión de los miembros de un hogar para hacerse conectar el internet o no, desde una red pública hacia su hogar.	Esta variable fue medida con el cuestionario de la Encuesta Nacional de Hogares (ENAH), a través del módulo 100.	Acceso a internet	1=Tiene conexión internet 0=No conexión internet	a tiene a <b>p1144:</b> Su hogar tiene: ¿Conexión a internet?
				Material de vivienda	1=Rústico 0=No rústico
<b>Variable 2:</b> <b>Factores de acceso a internet</b> (Variable independiente)	Los factores de acceso a internet son las covariables que explican el simple hecho de que los miembros de un hogar decidan o no hacerse instalar la conexión de internet disponible en su hogar, siempre en cuando la red o infraestructura de internet exista en el lugar donde este ubicada la vivienda.	Esta variable fue medida con el cuestionario de la Encuesta Nacional de Hogares (ENAH), a través de los módulos 100, 200, 300, 612 y 34.	Habitaciones	Número de habitaciones en el hogar	<b>p104:</b> ¿Cuántas habitaciones en total tiene la vivienda, sin contar el baño, la cocina, los pasadizos, ni el garaje?
			Energía eléctrica	1=Si 0=No	<b>p1121:</b> Tipo de alumbrado del hogar: electricidad.
			Miembros	Número de miembros en el hogar	<b>mieperho:</b> Total de miembros del hogar.
			Niños	1=Si 0=No	<b>p208a:</b> ¿Qué edad tiene en años cumplidos?
			Estudiantes	1=Si 0=No	<b>p303:</b> El año pasado (...), ¿estuvo matriculado en algún centro o programa de educación básica o superior? <b>p306:</b> Este año, ¿está matriculado en algún centro o programa de educación básica o superior?
			Celular	1=Si 0=No	<b>p1142:</b> Su hogar tiene: ¿teléfono celular?
			Computadora	1=Si 0=No	<b>p612n:</b> Equipamiento del hogar. <b>p612:</b> ¿Su hogar tiene?
			Región geográfica	1=Costa 2=Sierra 3=Selva	<b>dominio:</b> Dominio geográfico.
Zona de residencia	1=Zona Urbana 0=Zona rural	<b>estrato:</b> Estrato geográfico.			

### ANEXO 3. Matriz de operacionalización de variables para el O.E.2.

VARIABLES	DEFINICIÓN CONCEPTUAL	DEFINICIÓN OPERACIONAL	DIMENSIONES	INDICADORES	ITEMS
<b>Variable 3:</b> <b>Uso de internet</b> (Variable dependiente)	Es la disposición o decisión de un jefe de hogar para hacer el uso de internet o no, en cualquier lugar.	Esta variable fue medida con el cuestionario de la Encuesta Nacional de Hogares (ENAH), a través del módulo 300.	Uso de internet	1=Usa el internet 0=No usa el internet	<b>p314a:</b> En el mes anterior, ¿Ud. hizo uso del servicio de internet?
			Edad	Número de años cumplidos	<b>p208a:</b> ¿Qué edad tiene en años cumplidos? (en años).
			Genero	1=Hombre 0=Mujer	<b>p207:</b> Sexo.
			Estado civil	1=Casado 0=Soltero	<b>p209:</b> ¿Cuál es su estado civil o conyugal?
			Educación	1=Inicial 2=Primaria 3=Secundaria 4=Superior tecnológica 5=Superior universitaria 6=Maestría/doctorado	<b>p301a:</b> ¿Cuál es el último año o grado de estudios y nivel que aprobó? – nivel.
			Miembros	Número de miembros	<b>mieperho:</b> Total de miembros del hogar.
			Niños	1=Si 0=No	<b>p208a:</b> ¿Qué edad tiene en años cumplidos?
			Estudiantes	1=Si 0=No	<b>p303:</b> El año pasado (...), ¿estuvo matriculado en algún centro o programa de educación básica o superior? <b>p306:</b> Este año, ¿está matriculado en algún centro o programa de educación básica o superior?
			Celular	1=Si 0=No	<b>p1142:</b> su hogar tiene: ¿teléfono celular?
			Computadora	1=Si 0=No	<b>p612n:</b> Equipamiento del hogar. <b>p612:</b> ¿Su hogar tiene?
Región geográfica	1=Costa 2=Sierra 3=Selva	<b>dominio:</b> Dominio geográfico.			
Zona de residencia	1=Zona urbana 0=Zona rural	<b>estrato:</b> Estrato geográfico.			
<b>Variable 4:</b> <b>Factores de uso de internet</b> (Variable independiente)	Los factores de uso de internet son las covariables que explican el simple hecho de que un jefe de hogar decida o no usar el internet, en cualquier lugar.	Esta variable fue medida con el cuestionario de la Encuesta Nacional de Hogares (ENAH), a través de los módulos 200, 300, 612 y 34.			

#### ANEXO 4. Matriz de operacionalización de variables para el O.E.3.

VARIABLES	DEFINICIÓN CONCEPTUAL	DEFINICIÓN OPERACIONAL	DIMENSIONES	INDICADORES	ITEMS
<b>Variable Ingreso</b> (Variable resultado)	<b>5:</b> Es el dinero que recibe un agente económico debido a su actividad económica.	Esta variable fue medida con el cuestionario de la Encuesta Nacional de Hogares (ENAH), a través de los módulos 500 y 34.	Ingreso	Ingreso neto per cápita	<b>inghog2d:</b> Ingreso neto total. <b>ocu500:</b> Indicador de la pea.
<b>Variable 6: Uso de internet</b> (Variable tratamiento)	Es la disposición o decisión de un jefe de hogar en condición de PEA para hacer el uso de internet o no, desde cualquier lugar.	Esta variable fue medida con el cuestionario de la Encuesta Nacional de Hogares (ENAH), a través del módulo 300.	Uso de internet	1=Usa el internet 0=No usa el internet	<b>p314a:</b> En el mes anterior, ¿Ud. hizo uso del servicio de internet?
<b>Variable 7: Características observables</b> (Variable control)	Es la condición que determina el uso o no uso del internet, por parte del jefe de hogar, desde cualquier lugar.	Esta variable fue medida con el cuestionario de la Encuesta Nacional de Hogares (ENAH), a través de los módulos 200, 300, 612 y 34.	Edad	Número de años cumplidos	<b>p208a:</b> ¿Qué edad tiene en años cumplidos? (en años).
			Genero	1=Hombre 0=Mujer	<b>p207:</b> Sexo.
			Estado civil	1=Casado 0=Soltero	<b>p209:</b> ¿Cuál es su estado civil o conyugal?
			Educación	1=Inicial 2=Primaria 3=Secundaria 4=Superior tecnológica 5=Superior universitaria 6=Maestría/doctorado	<b>p301a:</b> ¿Cuál es el último año o grado de estudios y nivel que aprobó? – nivel.
			Miembros	Número de miembros	<b>mieperho:</b> Total de miembros del hogar.
			Niños	1=Si 0=No	<b>p208a:</b> ¿Qué edad tiene en años cumplidos?
			Estudiantes	1=Si 0=No	<b>p303:</b> El año pasado (...), ¿estuvo matriculado en algún centro o programa de educación básica o superior? <b>p306:</b> Este año, ¿está matriculado en algún centro o programa de educación básica o superior?
			Celular	1=Si 0=No	<b>p1142:</b> su hogar tiene: ¿teléfono celular?
			Computadora	1=Si 0=No	<b>p612n:</b> Equipamiento del hogar. <b>p612:</b> ¿Su hogar tiene?
			Región geográfica	1=Costa 2=Sierra 3=Selva	<b>dominio:</b> Dominio geográfico.
Zona de residencia	1=Zona urbana 0=Zona rural	<b>estrato:</b> Estrato geográfico.			



### ANEXO 5. Estimación del modelo inicial Logit O.E.1.

Iteration 0:	log likelihood =	-16935.339				
Iteration 1:	log likelihood =	-11097.819				
Iteration 2:	log likelihood =	-10796.459				
Iteration 3:	log likelihood =	-10769.586				
Iteration 4:	log likelihood =	-10768.102				
Iteration 5:	log likelihood =	-10768.1				
Iteration 6:	log likelihood =	-10768.1				
Logistic regression				Number of obs	=	26,663
				LR chi2(11)	=	12334.48
				Prob > chi2	=	0.0000
Log likelihood =	-10768.1			Pseudo R2	=	0.3642
acceso	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
material	-.6888366	.0402271	-17.12	0.000	-.7676802	-.6099929
habitaciones	.1798923	.0116343	15.46	0.000	.1570895	.202695
electricidad	.5843245	.1165709	5.01	0.000	.3558497	.8127993
miembros	-.0183131	.0137466	-1.33	0.183	-.0452559	.0086296
niños	.0193999	.0433942	0.45	0.655	-.065651	.1044509
estudiantes	.2943387	.0434916	6.77	0.000	.2090967	.3795807
celular	2.790453	.2575803	10.83	0.000	2.285605	3.295301
computadora	2.123889	.0372324	57.04	0.000	2.050915	2.196863
sierra	-.655897	.0421637	-15.56	0.000	-.7385364	-.5732576
selva	-.270277	.0478488	-5.65	0.000	-.364059	-.176495
urbano	1.00845	.0438133	23.02	0.000	.9225776	1.094322
_cons	-5.362416	.281589	-19.04	0.000	-5.91432	-4.810511

## ANEXO 6. Estimación del modelo final Logit O.E.1.

Iteration 0: log likelihood = -16935.339						
Iteration 1: log likelihood = -11098.79						
Iteration 2: log likelihood = -10797.355						
Iteration 3: log likelihood = -10770.495						
Iteration 4: log likelihood = -10769.014						
Iteration 5: log likelihood = -10769.01						
Iteration 6: log likelihood = -10769.01						
Logistic regression						
					Number of obs	= 26,663
					LR chi2(9)	= 12332.66
					Prob > chi2	= 0.0000
Log likelihood = -10769.01					Pseudo R2	= 0.3641
acceso	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
material	-.6924808	.0401265	-17.26	0.000	-.7711273	-.6138342
habitaciones	.1759605	.0112358	15.66	0.000	.1539386	.1979823
electricidad	.5904318	.1165065	5.07	0.000	.3620834	.8187803
estudiantes	.2623702	.0361895	7.25	0.000	.1914401	.3333003
celular	2.78394	.2574254	10.81	0.000	2.279395	3.288484
computadora	2.122985	.0371077	57.21	0.000	2.050255	2.195715
sierra	-.6532521	.0421057	-15.51	0.000	-.7357779	-.5707264
selva	-.2695107	.0478212	-5.64	0.000	-.3632385	-.1757828
urbano	1.010523	.043787	23.08	0.000	.9247026	1.096344
_cons	-5.388399	.2807842	-19.19	0.000	-5.938726	-4.838072

## ANEXO 7. Medidas de ajuste del modelo final Logit O.E.1.

Measures of Fit for logit of acceso			
Log-Lik Intercept Only:	-16935.339	Log-Lik Full Model:	-10769.010
D(26653):	21538.020	LR(9):	12332.659
		Prob > LR:	0.000
McFadden's R2:	0.364	McFadden's Adj R2:	0.364
ML (Cox-Snell) R2:	0.370	Cragg-Uhler(Nagelkerke) R2:	0.515
McKelvey & Zavoina's R2:	0.549	Efron's R2:	0.424
Variance of y*:	7.296	Variance of error:	3.290
Count R2:	0.825	Adj Count R2:	0.472
AIC:	0.809	AIC*n:	21558.020
BIC:	-250083.559	BIC':	-12240.939
BIC used by Stata:	21639.930	AIC used by Stata:	21558.020

### ANEXO 8. Estimación del modelo inicial Probit O.E.1.

Iteration 0:	log likelihood = -16935.339					
Iteration 1:	log likelihood = -10911.543					
Iteration 2:	log likelihood = -10773.064					
Iteration 3:	log likelihood = -10763.271					
Iteration 4:	log likelihood = -10763.029					
Iteration 5:	log likelihood = -10763.028					
Probit regression					Number of obs	= 26,663
					LR chi2(11)	= 12344.62
					Prob > chi2	= 0.0000
Log likelihood = -10763.028					Pseudo R2	= 0.3645
acceso	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
material	-.4063767	.0234623	-17.32	0.000	-.4523621	-.3603914
habitaciones	.0992991	.0065648	15.13	0.000	.0864323	.112166
electricidad	.2830324	.0572935	4.94	0.000	.1707393	.3953255
miembros	-.0099066	.0077475	-1.28	0.201	-.0250915	.0052782
niños	.0130528	.024467	0.53	0.594	-.0349016	.0610072
estudiantes	.1665431	.0245092	6.80	0.000	.118506	.2145802
celular	1.474508	.1300733	11.34	0.000	1.219569	1.729447
computadora	1.269528	.0217527	58.36	0.000	1.226894	1.312163
sierra	-.3704528	.0238205	-15.55	0.000	-.4171401	-.3237654
selva	-.161049	.0271717	-5.93	0.000	-.2143046	-.1077934
urbano	.5633537	.0243126	23.17	0.000	.515702	.6110055
_cons	-2.889335	.1427612	-20.24	0.000	-3.169142	-2.609528

### ANEXO 9. Estimación del modelo final Probit O.E.1.

Iteration 0: log likelihood = -16935.339						
Iteration 1: log likelihood = -10912.423						
Iteration 2: log likelihood = -10773.869						
Iteration 3: log likelihood = -10764.091						
Iteration 4: log likelihood = -10763.85						
Iteration 5: log likelihood = -10763.849						
Probit regression			Number of obs		= 26,663	
			LR chi2(9)		= 12342.98	
			Prob > chi2		= 0.0000	
Log likelihood = -10763.849			Pseudo R2		= 0.3644	
acceso	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
material	-.4083504	.0234014	-17.45	0.000	-.4542163	-.3624844
habitaciones	.0971595	.0063434	15.32	0.000	.0847266	.1095924
electricidad	.2857697	.057243	4.99	0.000	.1735755	.397964
estudiantes	.1497609	.0203856	7.35	0.000	.109806	.1897159
celular	1.471177	.1299407	11.32	0.000	1.216498	1.725856
computadora	1.268882	.0216872	58.51	0.000	1.226375	1.311388
sierra	-.3691362	.0237899	-15.52	0.000	-.4157635	-.3225089
selva	-.1605688	.0271574	-5.91	0.000	-.2137963	-.1073414
urbano	.5644951	.024297	23.23	0.000	.5168737	.6121164
_cons	-2.902424	.1422508	-20.40	0.000	-3.181231	-2.623618

### ANEXO 10. Medidas de ajuste del modelo final Probit O.E.1.

Measures of Fit for probit of acceso			
Log-Lik Intercept Only:	-16935.339	Log-Lik Full Model:	-10763.849
D(26653):	21527.699	LR(9):	12342.980
		Prob > LR:	0.000
McFadden's R2:	0.364	McFadden's Adj R2:	0.364
ML (Cox-Snell) R2:	0.371	Cragg-Uhler(Nagelkerke) R2:	0.515
McKelvey & Zavoina's R2:	0.565	Efron's R2:	0.424
Variance of y*:	2.298	Variance of error:	1.000
Count R2:	0.825	Adj Count R2:	0.472
AIC:	0.808	AIC*n:	21547.699
BIC:	-250093.880	BIC':	-12251.261
BIC used by Stata:	21629.609	AIC used by Stata:	21547.699

### ANEXO 11. Efectos marginales del modelo elegido O.E.1.

Marginal effects after probit								
y = Pr(acceso) (predict)								
= .25619444								
variable	dy/dx	Std. Err.	z	P> z	[	95% C.I.	]	X
material*	-.1380528	.00828	-16.67	0.000	-.154288	-.121818		.715373
habita~s	.0312751	.00205	15.24	0.000	.027254	.035296		3.24986
electr~d*	.0842288	.01516	5.56	0.000	.054515	.113942		.924202
estudi~s*	.0477161	.00642	7.43	0.000	.035124	.060308		.59802
celular*	.2662214	.00743	35.81	0.000	.251651	.280792		.936279
comput~a*	.4443609	.00765	58.06	0.000	.429359	.459363		.287815
sierra*	-.1159697	.00728	-15.93	0.000	-.130234	-.101705		.413982
selva*	-.0501261	.00821	-6.11	0.000	-.066218	-.034035		.223231
urbano*	.176394	.00726	24.30	0.000	.162169	.190619		.559839

(\*) dy/dx is for discrete change of dummy variable from 0 to 1

### ANEXO 12. Efectos marginales del modelo elegido (sierra=0 y selva=0) O.E.1.

Marginal effects after probit								
y = Pr(acceso) (predict)								
= .3204421								
variable	dy/dx	Std. Err.	z	P> z	[	95% C.I.	]	X
material*	-.1507483	.00871	-17.31	0.000	-.167819	-.133678		.715373
habita~s	.0347654	.00228	15.24	0.000	.030293	.039238		3.24986
electr~d*	.095711	.01765	5.42	0.000	.061122	.130301		.924202
estudi~s*	.0531775	.00719	7.40	0.000	.039085	.06727		.59802
celular*	.32208	.01096	29.39	0.000	.300603	.343557		.936279
comput~a*	.4662238	.00756	61.66	0.000	.451405	.481042		.287815
sierra*	-.1187519	.00766	-15.50	0.000	-.133771	-.103733		0
selva*	-.0551229	.00922	-5.98	0.000	-.073185	-.037061		0
urbano*	.1967532	.00784	25.08	0.000	.181379	.212127		.559839

(\*) dy/dx is for discrete change of dummy variable from 0 to 1



### ANEXO 13. Estimación del modelo inicial Logit O.E.2.

Iteration 0:	log likelihood =	-18403.19				
Iteration 1:	log likelihood =	-10522.558				
Iteration 2:	log likelihood =	-10238.435				
Iteration 3:	log likelihood =	-10219.883				
Iteration 4:	log likelihood =	-10219.749				
Iteration 5:	log likelihood =	-10219.749				
Logistic regression			Number of obs	=	26,663	
			LR chi2(16)	=	16366.88	
			Prob > chi2	=	0.0000	
Log likelihood = -10219.749			Pseudo R2	=	0.4447	
uso_int	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
edad	-.0749	.0017807	-42.06	0.000	-.07839	-.07141
genero	.2125593	.0527332	4.03	0.000	.1092042	.3159144
est_civil	-.2592908	.0534648	-4.85	0.000	-.3640798	-.1545017
primaria	1.523403	.2185511	6.97	0.000	1.09505	1.951755
secundaria	2.6669	.2177541	12.25	0.000	2.24011	3.09369
sup_no_univ	3.932262	.2234855	17.60	0.000	3.494238	4.370285
sup_univ	4.669683	.2299102	20.31	0.000	4.219067	5.120298
maestria	6.907507	.5042538	13.70	0.000	5.919187	7.895826
miembros	-.1181662	.0149839	-7.89	0.000	-.1475342	-.0887983
niños	.0700153	.0461952	1.52	0.130	-.0205256	.1605563
estudiantes	.1242103	.0475856	2.61	0.009	.0309441	.2174764
celular	2.655561	.1669798	15.90	0.000	2.328286	2.982835
computadora	1.028861	.0444056	23.17	0.000	.9418277	1.115894
sierra	-1.052029	.0436404	-24.11	0.000	-1.137563	-.9664958
selva	-.8861168	.047853	-18.52	0.000	-.979907	-.7923265
urbano	.9176846	.0408444	22.47	0.000	.837631	.9977383
_cons	-1.511205	.2882003	-5.24	0.000	-2.076067	-.9463424

#### ANEXO 14. Estimación del modelo final Logit O.E.2.

Iteration 0: log likelihood = -18403.19						
Iteration 1: log likelihood = -10524.795						
Iteration 2: log likelihood = -10239.687						
Iteration 3: log likelihood = -10221.033						
Iteration 4: log likelihood = -10220.898						
Iteration 5: log likelihood = -10220.897						
Logistic regression						
					Number of obs	= 26,663
					LR chi2(15)	= 16364.58
					Prob > chi2	= 0.0000
Log likelihood = -10220.897					Pseudo R2	= 0.4446
uso_int	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
edad	-.0757556	.0016916	-44.78	0.000	-.0790711	-.07244
genero	.2106094	.0527393	3.99	0.000	.1072422	.3139766
est_civil	-.2601177	.0534823	-4.86	0.000	-.3649411	-.1552943
primaria	1.523747	.2186427	6.97	0.000	1.095215	1.952279
secundaria	2.669301	.2178377	12.25	0.000	2.242347	3.096255
sup_no_univ	3.935456	.2235641	17.60	0.000	3.497278	4.373633
sup_univ	4.674473	.2299813	20.33	0.000	4.223718	5.125228
maestria	6.91318	.5043309	13.71	0.000	5.924709	7.90165
miembros	-.1076411	.0132623	-8.12	0.000	-.1336347	-.0816476
estudiantes	.1114368	.0468235	2.38	0.017	.0196645	.2032091
celular	2.651623	.1669503	15.88	0.000	2.324406	2.97884
computadora	1.022166	.0441781	23.14	0.000	.9355786	1.108753
sierra	-1.052244	.0436392	-24.11	0.000	-1.137775	-.9667124
selva	-.8841399	.0478286	-18.49	0.000	-.9778822	-.7903975
urbano	.9164509	.0408253	22.45	0.000	.8364347	.996467
_cons	-1.471602	.2870685	-5.13	0.000	-2.034245	-.9089577

#### ANEXO 15. Medidas de ajuste del modelo final Logit O.E.2.

Measures of Fit for logit of uso_int			
Log-Lik Intercept Only:	-18403.190	Log-Lik Full Model:	-10220.897
D(26647):	20441.795	LR(15):	16364.584
		Prob > LR:	0.000
McFadden's R2:	0.445	McFadden's Adj R2:	0.444
ML (Cox-Snell) R2:	0.459	Cragg-Uhler(Nagelkerke) R2:	0.613
McKelvey & Zavoina's R2:	0.696	Efron's R2:	0.504
Variance of y*:	10.824	Variance of error:	3.290
Count R2:	0.821	Adj Count R2:	0.612
AIC:	0.768	AIC*n:	20473.795
BIC:	-251118.638	BIC':	-16211.719
BIC used by Stata:	20604.851	AIC used by Stata:	20473.795



### ANEXO 16. Estimación del modelo inicial Probit O.E.2.

Iteration 0:	log likelihood =	-18403.19				
Iteration 1:	log likelihood =	-10514.16				
Iteration 2:	log likelihood =	-10233.302				
Iteration 3:	log likelihood =	-10222.098				
Iteration 4:	log likelihood =	-10222.043				
Iteration 5:	log likelihood =	-10222.043				
Probit regression			Number of obs	=	26,663	
			LR chi2(16)	=	16362.29	
			Prob > chi2	=	0.0000	
Log likelihood = -10222.043			Pseudo R2	=	0.4446	
uso_int	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
edad	-.04254	.0009879	-43.06	0.000	-.0444762	-.0406037
genero	.117831	.0301006	3.91	0.000	.058835	.176827
est_civil	-.1394827	.0304427	-4.58	0.000	-.1991493	-.079816
primaria	.7411105	.1028436	7.21	0.000	.5395408	.9426803
secundaria	1.416246	.1024668	13.82	0.000	1.215414	1.617077
sup_no_univ	2.132526	.106112	20.10	0.000	1.924551	2.340502
sup_univ	2.537405	.1096309	23.14	0.000	2.322533	2.752278
maestria	3.489836	.2025894	17.23	0.000	3.092768	3.886904
miembros	-.0670873	.0085004	-7.89	0.000	-.0837479	-.0504268
niños	.0373525	.0264626	1.41	0.158	-.0145133	.0892182
estudiantes	.0710748	.0271178	2.62	0.009	.017925	.1242247
celular	1.462734	.0871731	16.78	0.000	1.291877	1.63359
computadora	.5957445	.0254314	23.43	0.000	.5458998	.6455891
sierra	-.5992844	.0248594	-24.11	0.000	-.648008	-.5505609
selva	-.5038916	.0273713	-18.41	0.000	-.5575384	-.4502448
urbano	.5291679	.0235404	22.48	0.000	.4830296	.5753063
_cons	-.7176062	.1443347	-4.97	0.000	-1.000497	-.4347154

### ANEXO 17. Estimación del modelo final Probit O.E.2.

Iteration 0: log likelihood = -18403.19						
Iteration 1: log likelihood = -10516.574						
Iteration 2: log likelihood = -10234.436						
Iteration 3: log likelihood = -10223.096						
Iteration 4: log likelihood = -10223.039						
Iteration 5: log likelihood = -10223.039						
Probit regression			Number of obs		= 26,663	
			LR chi2(15)		= 16360.30	
			Prob > chi2		= 0.0000	
Log likelihood = -10223.039			Pseudo R2		= 0.4445	
uso_int	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
edad	-.0429924	.0009355	-45.96	0.000	-.0448259	-.0411589
genero	.1167373	.0300983	3.88	0.000	.0577457	.175729
est_civil	-.1397324	.0304512	-4.59	0.000	-.1994156	-.0800491
primaria	.7414196	.1029373	7.20	0.000	.5396662	.9431729
secundaria	1.417632	.1025544	13.82	0.000	1.216629	1.618635
sup_no_univ	2.134529	.106192	20.10	0.000	1.926397	2.342662
sup_univ	2.540027	.1097042	23.15	0.000	2.325011	2.755044
maestria	3.492582	.2026469	17.23	0.000	3.095401	3.889762
miembros	-.0615324	.007531	-8.17	0.000	-.0762929	-.046772
estudiantes	.0642657	.0266833	2.41	0.016	.0119675	.116564
celular	1.461045	.0871845	16.76	0.000	1.290166	1.631923
computadora	.5920581	.0252933	23.41	0.000	.5424841	.6416321
sierra	-.5994141	.0248596	-24.11	0.000	-.6481381	-.5506901
selva	-.5027687	.0273579	-18.38	0.000	-.5563891	-.4491483
urbano	.528753	.0235367	22.47	0.000	.4826219	.5748841
_cons	-.6971988	.1436815	-4.85	0.000	-.9788093	-.4155882

### ANEXO 18. Medidas de ajuste del modelo final Probit O.E.2.

Measures of Fit for probit of uso_int			
Log-Lik Intercept Only:	-18403.190	Log-Lik Full Model:	-10223.039
D(26647):	20446.078	LR(15):	16360.301
		Prob > LR:	0.000
McFadden's R2:	0.444	McFadden's Adj R2:	0.444
ML (Cox-Snell) R2:	0.459	Cragg-Uhler(Nagelkerke) R2:	0.613
McKelvey & Zavoina's R2:	0.704	Efron's R2:	0.504
Variance of y*:	3.374	Variance of error:	1.000
Count R2:	0.821	Adj Count R2:	0.613
AIC:	0.768	AIC*n:	20478.078
BIC:	-251114.355	BIC':	-16207.435
BIC used by Stata:	20609.135	AIC used by Stata:	20478.078



**ANEXO 19.** Efectos marginales del modelo elegido O.E.2.

Marginal effects after logit							
y = Pr(uso_int) (predict)							
= .40744324							
variable	dy/dx	Std. Err.	z	P> z	[	95% C.I.	]
edad	-.0182899	.00041	-44.78	0.000	-.01909	-.017489	50.3237
genero*	.0503023	.01244	4.04	0.000	.025913	.074692	.740427
est_ci~l*	-.0632029	.01305	-4.84	0.000	-.088788	-.037618	.669092
primaria*	.3621204	.04786	7.57	0.000	.268322	.455918	.341747
secund~a*	.5830943	.03606	16.17	0.000	.512414	.653775	.373101
sup_no~v*	.6533277	.01551	42.12	0.000	.622924	.683731	.115516
sup_univ*	.6770121	.01141	59.35	0.000	.654654	.699371	.099239
maestria*	.6261398	.00606	103.25	0.000	.614254	.638025	.021453
miembros	-.0259881	.00321	-8.10	0.000	-.032274	-.019702	3.52635
estudi~s*	.0268432	.01125	2.39	0.017	.004796	.04889	.59802
celular*	.394475	.01004	39.30	0.000	.374803	.414147	.936279
comput~a*	.2486691	.01049	23.70	0.000	.228102	.269237	.287815
sierra*	-.2445822	.00962	-25.42	0.000	-.263439	-.225726	.413982
selva*	-.1987633	.00985	-20.17	0.000	-.218075	-.179451	.223231
urbano*	.2156064	.00919	23.47	0.000	.197603	.233609	.559839

(\*) dy/dx is for discrete change of dummy variable from 0 to 1



**ANEXO 20.** Efectos marginales del modelo elegido (primaria=0, secundaria=0, sup\_no\_univ=0, sup\_univ=0 y maestria=0) O.E.2.

Marginal effects after logit							
y = Pr(uso_int) (predict)							
= .04935954							
variable	dy/dx	Std. Err.	z	P> z	[	95% C.I.	]
edad	-.0035547	.0007	-5.08	0.000	-.004925	-.002185	50.3237
genero*	.0094532	.00299	3.16	0.002	.003586	.015321	.740427
est_ci~l*	-.0127233	.00365	-3.48	0.000	-.019886	-.005561	.669092
primaria*	.1430766	.01144	12.51	0.000	.120661	.165492	0
secund~a*	.3789617	.012	31.57	0.000	.355437	.402486	0
sup_no~v*	.6772426	.01518	44.61	0.000	.64749	.706996	0
sup_univ*	.7983138	.01428	55.92	0.000	.770335	.826293	0
maestria*	.9318452	.01318	70.68	0.000	.906004	.957687	0
miembros	-.0050509	.00117	-4.33	0.000	-.007339	-.002763	3.52635
estudi~s*	.0051796	.00238	2.18	0.029	.000519	.009841	.59802
celular*	.0536014	.0109	4.92	0.000	.032239	.074964	.936279
comput~a*	.0598378	.01172	5.11	0.000	.036869	.082807	.287815
sierra*	-.0470415	.00933	-5.04	0.000	-.065336	-.028747	.413982
selva*	-.0340267	.00689	-4.94	0.000	-.04754	-.020514	.223231
urbano*	.0419701	.00838	5.01	0.000	.025543	.058397	.559839

(\*) dy/dx is for discrete change of dummy variable from 0 to 1

**ANEXO 21.** Efectos marginales del modelo elegido (sierra=0 y selva=0) O.E.2.

Marginal effects after logit							
y = Pr(uso_int) (predict)							
= .56425081							
variable	dy/dx	Std. Err.	z	P> z	[ 95% C.I. ]	X	
edad	-.0186262	.00042	-44.20	0.000	-.019452 - .0178	50.3237	
genero*	.0520415	.01308	3.98	0.000	.026404 .077679	.740427	
est_ci~l*	-.0634815	.01295	-4.90	0.000	-.088861 -.038102	.669092	
primaria*	.344477	.04382	7.86	0.000	.258587 .430367	.341747	
secund~a*	.5498962	.03409	16.13	0.000	.483081 .616712	.373101	
sup_no~v*	.525668	.01478	35.56	0.000	.496691 .554645	.115516	
sup_univ*	.5398554	.01251	43.17	0.000	.515344 .564367	.099239	
maestria*	.4716067	.00879	53.65	0.000	.454379 .488834	.021453	
miembros	-.0264659	.00326	-8.13	0.000	-.032848 -.020084	3.52635	
estudi~s*	.0274301	.01153	2.38	0.017	.004823 .050037	.59802	
celular*	.5076559	.01592	31.89	0.000	.476456 .538856	.936279	
comput~a*	.2373166	.00953	24.91	0.000	.218645 .255988	.287815	
sierra*	-.2529	.01011	-25.02	0.000	-.27271 -.23309	0	
selva*	-.2157664	.01124	-19.19	0.000	-.237801 -.193732	0	
urbano*	.2229851	.00989	22.54	0.000	.203594 .242376	.559839	

(\*) dy/dx is for discrete change of dummy variable from 0 to 1



**ANEXO 22.** Estimación de la puntuación de propensión O.E.3.

Logistic regression		Number of obs	=	26,483		
		LR chi2(15)	=	16232.04		
		Prob > chi2	=	0.0000		
Log likelihood = -10169.257		Pseudo R2	=	0.4439		
uso_int	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
edad	-.0758345	.0017006	-44.59	0.000	-.0791676	-.0725013
genero	.2178603	.0530734	4.10	0.000	.1138383	.3218823
est_civil	-.26519	.0537293	-4.94	0.000	-.3704976	-.1598824
primaria	1.508587	.2189271	6.89	0.000	1.079498	1.937677
secundaria	2.652116	.2181093	12.16	0.000	2.224629	3.079602
sup_no_univ	3.91525	.2238661	17.49	0.000	3.47648	4.354019
sup_univ	4.646707	.2302872	20.18	0.000	4.195352	5.098062
maestria	6.890878	.5044417	13.66	0.000	5.90219	7.879566
miembros	-.1081987	.013293	-8.14	0.000	-.1342525	-.0821448
estudiantes	.1131603	.0469472	2.41	0.016	.0211455	.2051751
celular	2.650059	.1669971	15.87	0.000	2.322751	2.977368
computadora	1.02763	.0443163	23.19	0.000	.9407712	1.114488
sierra	-1.052973	.0437643	-24.06	0.000	-1.138749	-.9671963
selva	-.8884643	.0479725	-18.52	0.000	-.9824886	-.79444
urbano	.9194381	.0409634	22.45	0.000	.8391512	.9997249
_cons	-1.451163	.2873711	-5.05	0.000	-2.0144	-.8879261

### ANEXO 23. Estimación del impacto (vecino más cercano) O.E.3.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
ln_ing_perc	Unmatched	6.65313861	6.01789385	.635244766	.009501637	66.86
	ATT	6.64961715	6.29424495	.355372199	.030141557	11.79

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support		Total
	Off suppo	On suppor	
Untreated	0	14,213	14,213
Treated	37	12,233	12,270
Total	37	26,446	26,483

### ANEXO 24. Prueba de la propiedad de balanceo (vecino más cercano) O.E.3.

Variable	Mean		%bias	t-test		V(T)/ V(C)
	Treated	Control		t	p> t	
edad	44.969	43.672	10.2	8.67	0.000	0.97
genero	.73629	.76987	-7.7	-6.09	0.000	.
est_civil	.65846	.66415	-1.2	-0.94	0.347	.
primaria	.11068	.11103	-0.1	-0.09	0.932	.
secundaria	.43562	.44267	-1.5	-1.11	0.267	.
sup_no_univ	.21254	.23097	-5.8	-3.47	0.001	.
sup_univ	.19586	.19254	1.1	0.66	0.512	.
maestria	.04333	.02099	15.0	9.92	0.000	.
miembros	3.5788	3.6662	-4.9	-4.27	0.000	1.08*
estudiantes	.67097	.70635	-7.3	-5.98	0.000	.
celular	.99583	.99666	-0.4	-1.07	0.286	.
computadora	.48917	.4508	9.2	6.02	0.000	.
sierra	.28513	.26214	4.8	4.03	0.000	.
selva	.18426	.18072	0.9	0.72	0.475	.
urbano	.79702	.78344	3.1	2.61	0.009	.

\* if variance ratio outside [0.97; 1.04]

Ps	R2	LR	chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0.009		317.41		0.000	4.9	4.8	22.8	1.41	50

\* if B>25%, R outside [0.5; 2]

### ANEXO 25. Estimación del impacto (radio y calibre) O.E.3.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
ln_ing_perc	Unmatched	6.65313861	6.01789385	.635244766	.009501637	66.86
	ATT	6.64961715	6.26956166	.380055491	.019312143	19.68

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support		Total
	Off suppo	On suppor	
Untreated	0	14,213	14,213
Treated	37	12,233	12,270
Total	37	26,446	26,483

### ANEXO 26. Prueba de la propiedad de balanceo (radio y calibre) O.E.3.

Variable	Mean		%bias	t-test		V(T)/ V(C)
	Treated	Control		t	p> t	
edad	44.969	45.444	-3.7	-3.19	0.001	0.98
genero	.73629	.75102	-3.4	-2.64	0.008	.
est_civil	.65846	.66948	-2.3	-1.82	0.068	.
primaria	.11068	.11786	-1.7	-1.76	0.078	.
secundaria	.43562	.49132	-11.6	-8.75	0.000	.
sup_no_univ	.21254	.24559	-10.5	-6.16	0.000	.
sup_univ	.19586	.13448	20.8	12.97	0.000	.
maestria	.04333	.00638	24.8	18.69	0.000	.
miembros	3.5788	3.6895	-6.2	-5.40	0.000	1.07*
estudiantes	.67097	.69071	-4.1	-3.31	0.001	.
celular	.99583	.99324	1.1	2.75	0.006	.
computadora	.48917	.42137	16.2	10.67	0.000	.
sierra	.28513	.28907	-0.8	-0.68	0.496	.
selva	.18426	.17803	1.5	1.26	0.206	.
urbano	.79702	.77719	4.5	3.79	0.000	.

\* if variance ratio outside [0.97; 1.04]

Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0.022	759.76	0.000	7.5	4.1	34.5*	3.03*	50

\* if B>25%, R outside [0.5; 2]

### ANEXO 27. Estimación del impacto (kernel) O.E.3.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
ln_ing_perc	Unmatched	6.65313861	6.01789385	.635244766	.009501637	66.86
	ATT	6.64961715	6.28737821	.362238938	.022256789	16.28

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support		Total
	Off suppo	On suppor	
Untreated	0	14,213	14,213
Treated	37	12,233	12,270
Total	37	26,446	26,483

### ANEXO 28. Prueba de la propiedad de balanceo (kernel) O.E.3.

Variable	Mean		%bias	t-test		V(T)/ V(C)
	Treated	Control		t	p> t	
edad	44.969	44.864	0.8	0.71	0.478	1.00
genero	.73629	.75594	-4.5	-3.53	0.000	.
est_civil	.65846	.66836	-2.1	-1.64	0.101	.
primaria	.11068	.11337	-0.6	-0.67	0.506	.
secundaria	.43562	.44833	-2.6	-2.00	0.046	.
sup_no_univ	.21254	.26275	-15.9	-9.24	0.000	.
sup_univ	.19586	.16198	11.5	6.92	0.000	.
maestria	.04333	.01107	21.6	15.58	0.000	.
miembros	3.5788	3.685	-5.9	-5.22	0.000	1.10*
estudiantes	.67097	.69383	-4.7	-3.84	0.000	.
celular	.99583	.99564	0.1	0.23	0.815	.
computadora	.48917	.4426	11.1	7.31	0.000	.
sierra	.28513	.27301	2.5	2.11	0.035	.
selva	.18426	.17166	3.0	2.58	0.010	.
urbano	.79702	.78552	2.6	2.21	0.027	.

\* if variance ratio outside [0.97; 1.04]

Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0.013	456.52	0.000	6.0	3.0	27.0*	2.24*	50

\* if B>25%, R outside [0.5; 2]



## ANEXO 29. Declaración jurada de autenticidad de tesis



Universidad Nacional  
del Altiplano Puno



VRI  
Vicerrectorado  
de Investigación



Repositorio  
Institucional

### DECLARACIÓN JURADA DE AUTENTICIDAD DE TESIS

Por el presente documento, Yo Bill Clinton Cutipa Mamani,  
identificado con DNI 70461394 en mi condición de egresado de:

Escuela Profesional,  Programa de Segunda Especialidad,  Programa de Maestría o Doctorado  
Ingeniería Económica

informo que he elaborado el/la  Tesis o  Trabajo de Investigación denominada:  
" Factores que inciden en el acceso y uso del internet y su  
impacto sobre la generación de empleo en los hogares del  
Perú, año 2020 "

Es un tema original.

Declaro que el presente trabajo de tesis es elaborado por mi persona y **no existe plagio/copia** de ninguna naturaleza, en especial de otro documento de investigación (tesis, revista, texto, congreso, o similar) presentado por persona natural o jurídica alguna ante instituciones académicas, profesionales, de investigación o similares, en el país o en el extranjero.

Dejo constancia que las citas de otros autores han sido debidamente identificadas en el trabajo de investigación, por lo que no asumiré como tuyas las opiniones vertidas por terceros, ya sea de fuentes encontradas en medios escritos, digitales o Internet.

Asimismo, ratifico que soy plenamente consciente de todo el contenido de la tesis y asumo la responsabilidad de cualquier error u omisión en el documento, así como de las connotaciones éticas y legales involucradas.

En caso de incumplimiento de esta declaración, me someto a las disposiciones legales vigentes y a las sanciones correspondientes de igual forma me someto a las sanciones establecidas en las Directivas y otras normas internas, así como las que me alcancen del Código Civil y Normas Legales conexas por el incumplimiento del presente compromiso

Puno 27 de septiembre del 2024

  
FIRMA (obligatoria)



Huella



## ANEXO 30. Autorización para el depósito de tesis en el Repositorio Institucional



Universidad Nacional  
del Altiplano Puno



Vicerrectorado  
de Investigación



Repositorio  
Institucional

### AUTORIZACIÓN PARA EL DEPÓSITO DE TESIS O TRABAJO DE INVESTIGACIÓN EN EL REPOSITORIO INSTITUCIONAL

Por el presente documento, Yo Bill Clinton Cotipa Mamani  
identificado con DNI 70461394 en mi condición de egresado de:

Escuela Profesional,  Programa de Segunda Especialidad,  Programa de Maestría o Doctorado

Ingeniería Económica  
informo que he elaborado el/la  Tesis o  Trabajo de Investigación denominada:

"Factores que inciden en el acceso y uso del internet y su  
impacto sobre la generación de empleo en los hogares del  
Perú, año 2020"

para la obtención de  Grado,  Título Profesional o  Segunda Especialidad.

Por medio del presente documento, afirmo y garantizo ser el legítimo, único y exclusivo titular de todos los derechos de propiedad intelectual sobre los documentos arriba mencionados, las obras, los contenidos, los productos y/o las creaciones en general (en adelante, los "Contenidos") que serán incluidos en el repositorio institucional de la Universidad Nacional del Altiplano de Puno.

También, doy seguridad de que los contenidos entregados se encuentran libres de toda contraseña, restricción o medida tecnológica de protección, con la finalidad de permitir que se puedan leer, descargar, reproducir, distribuir, imprimir, buscar y enlazar los textos completos, sin limitación alguna.

Autorizo a la Universidad Nacional del Altiplano de Puno a publicar los Contenidos en el Repositorio Institucional y, en consecuencia, en el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto, sobre la base de lo establecido en la Ley N° 30035, sus normas reglamentarias, modificatorias, sustitutorias y conexas, y de acuerdo con las políticas de acceso abierto que la Universidad aplique en relación con sus Repositorios Institucionales. Autorizo expresamente toda consulta y uso de los Contenidos, por parte de cualquier persona, por el tiempo de duración de los derechos patrimoniales de autor y derechos conexos, a título gratuito y a nivel mundial.

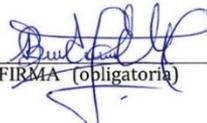
En consecuencia, la Universidad tendrá la posibilidad de divulgar y difundir los Contenidos, de manera total o parcial, sin limitación alguna y sin derecho a pago de contraprestación, remuneración ni regalía alguna a favor mío; en los medios, canales y plataformas que la Universidad y/o el Estado de la República del Perú determinen, a nivel mundial, sin restricción geográfica alguna y de manera indefinida, pudiendo crear y/o extraer los metadatos sobre los Contenidos, e incluir los Contenidos en los índices y buscadores que estimen necesarios para promover su difusión.

Autorizo que los Contenidos sean puestos a disposición del público a través de la siguiente licencia:

Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional. Para ver una copia de esta licencia, visita: <https://creativecommons.org/licenses/by-nc-sa/4.0/>

En señal de conformidad, suscribo el presente documento.

Puno 27 de septiembre del 2024

  
FIRMA (obligatoria)



Huella