

UNIVERSIDAD DE COSTA RICA
SISTEMA DE ESTUDIOS DE POSGRADO

CARACTERIZACIÓN DE LOS USUARIOS A PARTIR DE ESCALAS DE USO DE
INTERNET

EVALUACIÓN DEL TAMAÑO DE MUESTRA DE LA ENCUESTA DE ACCESO Y USO
DE LOS SERVICIOS DE TELECOMUNICACIONES EN COSTA RICA

Trabajo final de investigación aplicada sometida a la consideración de la Comisión del Programa
de Estudios de Posgrado en Estadística para optar por el grado y título de Maestría Profesional en
Estadística

VIVIAN AGUILAR AGUILAR

Ciudad Universitaria Rodrigo Facio, Costa Rica

2023

DEDICATORIA

A mi mamá por el sacrificio y apoyo en todo momento y a mis profesores por su guía y acompañamiento en el crecimiento personal y profesional.

AGRADECIMIENTO

Agradezco al Ministerio de Ciencia Tecnología y Telecomunicaciones por el interés mostrado en la temática de estudio, especialmente al Departamento de Análisis Económico y Mercados de Telecomunicaciones por el apoyo.

Mi agradecimiento y gratitud a mi profesor tutor de la Práctica Profesional I, Agustín Gómez Meléndez por el interés y apoyo; a mis lectores: Fernando Ramírez Hernández por sus aportes y Johnny Madrigal Pana por su guía y consejos.

Mi agradecimiento y gratitud a mi profesor tutor de la Práctica Profesional II, Johnny Madrigal Pana, por sus consejos, su apoyo y aportes; a mis lectores Fernando Ramírez Hernández, por el interés mostrado en el proyecto y Ricardo Alvarado Barrantes, por su apoyo, guía e inducción en el tema.

Este trabajo final de investigación aplicada I fue aceptado por la Comisión del Programa de Estudios de Posgrado en Estadística de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Maestría Profesional en Estadística.

M.Sc. Agustín Gómez Meléndez

Profesor Guía

M.Sc. Johnny Madrigal Pana

Lector

M.Sc. Fernando Ramírez Hernández

Lector

Vivian Aguilar Aguilar

Sustentante

Este trabajo final de investigación aplicada II fue aceptado por la Comisión del Programa de Estudios de Posgrado en Estadística de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Maestría Profesional en Estadística.

M.Sc. Johnny Madrigal Pana

Profesor Guía

M.Sc. Fernando Ramírez Hernández

Lector

Dr. Ricardo Alvarado Barrantes

Lector

Vivian Aguilar Aguilar

Sustentante

TABLA DE CONTENIDO

DEDICATORIA.....	ii
AGRADECIMIENTO.....	iii
TABLA DE CONTENIDO.....	vi
RESUMEN PRÁCTICA PROFESIONAL I.....	viii
RESUMEN PRÁCTICA PROFESIONAL II.....	ix
LISTA DE CUADROS.....	x
LISTA DE GRÁFICOS.....	x
LISTA DE FIGURAS.....	x
LISTA DE MAPAS.....	xi
LISTA DE ABREVIATURAS.....	xii
PRÁCTICA PROFESIONAL I.....	1
I. INTRODUCCIÓN.....	1
II. OBJETIVOS.....	3
2.1 Objetivo general.....	3
2.2 Objetivos específicos.....	4
III. ORIENTACIÓN CONCEPTUAL.....	4
IV. METODOLOGÍA.....	7
4.1 Fuente de datos.....	7
4.2 Variables de estudio.....	8
4.3 Técnicas estadísticas.....	11
4.4 Paquetes estadísticos.....	17
V. RESULTADOS.....	17
5.1 Construcción de parcelas y análisis factorial exploratorio.....	17
5.2 Análisis factorial confirmatorio.....	19
5.3 Construcción del índice y escalas.....	24
5.4 Análisis de conglomerados.....	31
VI. CONCLUSIONES.....	34
VII. BIBLIOGRAFÍA.....	38
VIII. ANEXOS.....	40

PRÁCTICA PROFESIONAL II	46
I. INTRODUCCIÓN	46
II. OBJETIVOS.....	48
2.1 Objetivo general	48
2.2 Objetivos específicos.....	48
III. ORIENTACIÓN CONCEPTUAL	48
IV. METODOLOGÍA	50
4.1 Fuente de datos	50
4.2 Variables de estudio.....	51
4.3 Técnicas estadísticas.....	55
4.3.1 Evaluación de la cantidad de UPM	56
4.3.2 Evaluación del tamaño de la submuestra	57
4.3.3 Resumen de escenarios desarrollados en el proceso de simulación.....	58
4.4 Software estadístico	60
V. RESULTADOS	60
5.1 Número de UPM en la muestra	61
5.2 Cantidad de entrevistas en cada UPM	67
VI. CONCLUSIONES	72
VII. BIBLIOGRAFÍA.....	75
VIII. ANEXOS	76

RESUMEN PRÁCTICA PROFESIONAL I

La apertura del mercado del sector de las telecomunicaciones en el 2008, trajo como consecuencia un aumento en el número de proveedores de estos servicios, lo cual contribuyó con el acceso de más personas a las TIC, es por esto que es indispensable la medición y monitoreo del sector. En el presente estudio se categorizan los diferentes tipos de usuarios de Internet por medio de la creación de escalas de uso utilizando los datos de la “Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017”. En primer lugar, se construyeron parcelas sustentadas con un análisis factorial exploratorio y un análisis de fiabilidad, se procedió con la construcción de un índice, las escalas de uso de Internet y se validó la teoría por medio de un análisis factorial confirmatorio, por último; se realizó un análisis de conglomerados para crear perfiles. Se crean cinco escalas: informativa, interacción social, entretenimiento, productividad, y habilitación ciudadana. Dentro de las escalas se pueden encontrar brechas por zona de residencia, sexo, grupos de edad, nivel educativo y de ingresos. Se determinan dos perfiles de usuarios, los consumidores sociales los cuales son personas que residen fuera de la región Central; una mayor proporción corresponde a mujeres, a personas de más de 45 años, personas con nivel educativo de primaria o menos, así como personas con ingresos inferiores a los 250 mil colones. Los ciudadanos digitales que en su mayoría son de la región Central, mayoritariamente hombres, con edades entre los 18 y 44 años, personas con nivel de instrucción universitaria y con ingresos superiores a los 750 mil colones.

RESUMEN PRÁCTICA PROFESIONAL II

El Ministerio de Ciencia, Innovación, Tecnología y Telecomunicaciones (MICITT), genera información actualizada y periódica sobre la evolución de las Tecnologías de Información y Comunicación (TIC) en las viviendas y población del país por medio de la “Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica (EAU)”. Ante la situación fiscal que atraviesa el país y las medidas de austeridad y contención del gasto público es necesario revisar el tamaño de muestra de la EAU y determinar un posible ajuste que, manteniendo la precisión en las estimaciones, reduzca los costos de la encuesta. Para lograrlo, se desarrolla un proceso de simulación en dos etapas: en la primera se determina el número de UPM y, en la segunda, la cantidad de entrevistas en cada UPM. Estos cálculos utilizan las variables más importantes de la encuesta, como lo son el índice y las escalas de uso de Internet. Para mantener la precisión y estabilidad en las estimaciones se establece que la diferencia entre el percentil 97,5 y 2,5 no sobrepase la cantidad de 0,5 en cada una de las regiones del país. Como resultado concluye que, en las seis regiones del país, se requiere 141 UPM, por lo que se logra una reducción del 29% (203 UPM en la muestra original). Con ese número de UPM y con un tamaño de submuestra aproximado de 18 entrevistas por UPM, la muestra final pasa de 3 618 a 2 556 entrevistas.

LISTA DE CUADROS

Cuadro 1. Distribución porcentual de las variables sociodemográficas de la EAU 2017	8
Cuadro 2. Estadísticos de bondad de ajuste y criterios en la evaluación SEM.....	16
Cuadro 3. Cargas factoriales y análisis de fiabilidad según parcelas	18
Cuadro 4. Estimación del modelo de uso de Internet, relación entre parcelas y variables	22
Cuadro 5. Estimación del modelo de uso de Internet, relación entre el constructo y las parcelas	23
Cuadro 6. Estadísticos de ajuste del AFC	23
Cuadro 7. Media e intervalos de confianza del índice de uso de Internet según variables sociodemográficas	26
Cuadro 8. Media e intervalos de confianza de las escalas de uso de Internet.....	27
Cuadro 9. Distribución porcentual de las personas entrevistadas por conglomerado según variables sociodemográficas	33
Cuadro 10. Distribución porcentual de las variables sociodemográficas	51
Cuadro 11. Media e intervalos de confianza del índice y las escalas de uso de Internet en el país según región de planificación.....	53
Cuadro 12. Distribución del tamaño de muestra y cantidad de UPM según región de planificación	58
Cuadro 13. Procesos de simulación utilizados para evaluar el tamaño de muestra de la EAU en cada una de las variables	59
Cuadro 14. Distribución del tamaño de muestra y cantidad de UPM al evaluar las estimaciones del índice de uso de Internet según región de planificación	65
Cuadro 15. Distribución la cantidad de UPM al evaluar las estimaciones de las escalas de uso de Internet según región de planificación	66
Cuadro 16. Cantidad de UPM de la muestra original y cantidad de UPM óptimo.....	67
Cuadro 17. Distribución de la muestra original, con UPM simuladas y con submuestra simuladas, según región de planificación, utilizando las estimaciones del índice de uso de Internet	72
Cuadro 18. Propuesta del nuevo tamaño de muestra para la EAU	74

LISTA DE GRÁFICOS

Gráfico 1. Uso de Internet según actividad	10
---	----

LISTA DE FIGURAS

Figura 1. Etapas empleadas en el análisis y técnicas estadísticas utilizadas.....	11
Figura 2. Proceso para realizar el planteamiento matemático en un modelo de ecuaciones estructurales.....	15
Figura 3. Planteamiento del modelo a estimar en el AFC.....	20
Figura 4. Estimación del modelo de uso de Internet por medio del diagrama de relaciones.....	21

Figura 5. Comportamiento de las medias del índice de uso de Internet según número de UPM simuladas en cada región de planificación	62
Figura 6. Diferencia entre los cuantiles de las medias del índice de uso de Internet según cantidad de UPM simuladas en cada región de planificación	63
Figura 7. Diferencia entre los cuantiles de las medias del índice de uso de Internet utilizando la reducción en la cantidad de UPM simuladas según tamaño de submuestra	68
Figura 8. Comportamiento de las medias del índice de uso de Internet según cantidad de submuestra simulada en cada región de planificación	69
Figura 9. Diferencia entre los cuantiles de las medias del índice de uso de Internet utilizando la cantidad de UPM de la muestra original según tamaño de submuestra	70

LISTA DE MAPAS

Mapa 1. Índice de uso de Internet según región de planificación	25
--	----

LISTA DE ABREVIATURAS

AFC	Análisis Factorial Confirmatorio
AFE	Análisis Factorial Exploratorio
EAU	Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica
ENAHO	Encuesta Nacional de Hogares
INEC	Instituto Nacional de Estadística y Censos
MICITT	Ministerio de Ciencia Innovación, Tecnología y Telecomunicaciones
OCDE	Organización para la Cooperación y el Desarrollo Económico
SUTEL	Superintendencia de Telecomunicaciones
TIC	Tecnologías de la Información y Comunicación
UCR	Universidad de Costa Rica
UIT	Unión Internacional de Telecomunicaciones
UPM	Unidad Primaria de Muestreo
USES	Unidad de Servicios Estadísticos
SEM	Modelo de Ecuaciones Estructurales

PRÁCTICA PROFESIONAL I

I. INTRODUCCIÓN

En Costa Rica, según lo señala la Superintendencia de Telecomunicaciones (SUTEL), la apertura del mercado del sector de las telecomunicaciones en el 2008, trajo como consecuencia un aumento en el número de proveedores de estos servicios, lo cual contribuyó con el acceso de más personas a las Tecnologías de la Información y Comunicación (TIC) (SUTEL, 2017).

Esta dinámica ha brindado un impulso al desarrollo social y económico, y por ende a la productividad y el conocimiento, construyendo así una economía globalizada y dando paso a una sociedad cada vez más conectada, lo que hace necesario la medición y monitoreo del sector.

Por un lado; según datos de la SUTEL la cantidad de suscripciones de telefonía móvil por cada 100 habitantes pasó de 69 en el 2010 a 169 en el 2019, lo que representa un incremento promedio anual de 11 %, dando como resultado un incremento total en los siete años de 145 %. Ligado al incremento en la cantidad de suscripciones de telefonía móvil se presenta un incremento de 669 % en la cantidad de suscripciones a Internet móvil¹ entre el 2010 y el 2019, pasando de 13 suscripciones por cada 100 habitantes a 92. Por otro lado; la cantidad de suscripciones a Internet fijo² creció de 137 % entre el 2010 y el 2019, se pasó de una penetración de 8 a 18 suscripciones por cada 100 habitantes.

¹ Las suscripciones a Internet móvil se refieren al abonado a redes celulares móviles (p. ej., empresas, personas) que pagan por el acceso a la Red de Internet (conexión TCP/IP)

² Las suscripciones a Internet fijo se refieren al abonado (p. ej., empresas, personas) que pagan por el acceso a la Red de Internet Fija (conexión TCP/IP)

De acuerdo con datos de la Encuesta Nacional de Hogares (ENAHOG), la penetración en el servicio de Internet en viviendas presenta un comportamiento creciente, para el 2010, el 26 % de las viviendas poseían acceso a Internet, ya sea por medio fijo o móvil y para el 2019 este porcentaje fue de 86 %. A esto se le suma un incremento en la cantidad de usuarios de Internet, el cual pasó de 34 % en 2010 a 81 % en el 2019, lo que representa un aumento de la cantidad de usuarios de Internet de 139 %.

Por lo tanto, debido al incremento en las conexiones a Internet, tanto fijas como móviles, así como el aumento acelerado en la cantidad de usuarios, es necesario evaluar el uso específico que le da la población al Internet, esto es parte del trabajo que realiza el Ministerio de Ciencia Tecnología y Telecomunicaciones (MICITT), con el fin de generar información actualizada y periódica sobre la evolución de los servicios de telecomunicaciones en las viviendas y en la población costarricense, con la “Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica”.

La encuesta se realizó anualmente desde el 2012 y hasta el 2018. Año tras año se han efectuado cambios en el contenido del cuestionario, buscando adaptar el instrumento para extraer indicadores actualizados en materia de telecomunicaciones y utilizando los estándares establecidos por la Organización para la Cooperación y el Desarrollo Económico (OCDE) y por la Unión Internacional de Telecomunicaciones (UIT).

Para el presente estudio se trabaja con los datos de la encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017, encuesta que realiza el MICITT. Esta brinda información sobre el acceso y uso de los servicios de telefonía fija, telefonía móvil, Internet fija, Internet móvil y televisión por suscripción. La encuesta incluye una batería de 33 preguntas sobre el uso de Internet en general, tanto fijo como móvil, esta batería está alineada con los conceptos dispuestos por la OCDE, por lo tanto, los indicadores planteados son solicitados por este organismo a los países miembro. Sin embargo; a la fecha los datos recopilados se utilizan de manera desagregada y descriptiva por parte de la OCDE, es por

esto que la creación de las escalas que se proponen en este documento permitiría simplificar la interpretación de los datos y brindar conclusiones más acertadas.

Actualmente, no existe ninguna iniciativa similar para la operacionalización de estos indicadores lo cual incrementa su aporte ya que la metodología empleada puede servir como insumo para la construcción de estos en otros países y eventualmente obtener datos comparativos en el nivel internacional con la creación de un índice global.

El propósito de las escalas es identificar el uso de Internet en la población costarricense, con el fin de caracterizar de una manera global el uso de Internet en Costa Rica, así como la creación de perfiles de usuarios que permita desagregar y evaluar con más detalle la información y así brindar resultados más robustos para la orientación de las políticas públicas en el sector.

Los resultados de esta investigación permitirán diseñar instrumentos dirigidos a los grupos con mayores desventajas en términos de alfabetización digital, considerando que el país ya superó la primera etapa de apertura del sector y se encuentra actualmente en un proceso de consolidación del mercado y apropiación de las TIC.

En el presente documento, inicialmente, se plantean los objetivos de la investigación, seguido de un apartado donde se expone la orientación conceptual en la cual se incluye una revisión de literatura sobre trabajos similares. Dentro de la metodología se define la fuente de datos, las variables de estudio, las técnicas estadísticas de análisis y los paquetes estadísticos utilizados. Seguidamente, se presentan los resultados y su análisis, finalizando con una sección de conclusiones y bibliografía utilizada.

II. OBJETIVOS

2.1 Objetivo general

Identificar y categorizar los diferentes tipos de usuarios de Internet por medio de la creación de escalas de uso en Costa Rica, a partir de la *Encuesta Acceso y uso de los servicios de*

Telecomunicaciones de 2017, con el propósito de aportar información para fundamentar la orientación de las políticas públicas en el sector.

2.2 Objetivos específicos

1. Caracterizar en el nivel nacional los usuarios de Internet de acuerdo con el índice y las escalas de uso.
2. Crear perfiles de usuarios a partir de las escalas de uso de Internet según región de planificación.

III. ORIENTACIÓN CONCEPTUAL

La creación de las escalas de uso de Internet y el índice se ha planteado a partir de un enfoque teórico, los autores han planteado diferentes perfiles de usuarios utilizando diversas metodologías para el análisis.

Torres e Infante (2011) realizaron un estudio a estudiantes universitarios y los clasificaron de acuerdo con la utilización e intensidad de uso de Internet. Por medio de un análisis factorial exploratorio obtienen ocho componentes, dentro de los cuales se tiene un componente que agrupa variables correspondientes a descargas de videos, música, programas y *software* en general.

En otro se vinculan variables de compra y venta en línea, variables de ocio como: ver televisión, escuchar radio y jugar en línea. Otro de los componentes agrupa variables referentes al uso de Internet en actividades académicas como *blogs*, uso de enciclopedias en línea para búsqueda de información, cursos o actividades académicas. Uno de los componentes corresponde al uso en redes sociales, mensajería de texto, correo electrónico.

Partiendo de los componentes se realiza un análisis de conglomerados y se crean tres perfiles de usuarios de Internet (perfil alto, perfil medio y perfil bajo). Se concluye que el nivel de ingresos incide en los perfiles de uso de Internet, los estudiantes con mayor poder adquisitivo

se encuentran en su mayoría en el perfil alto y por consiguiente los estudiantes con menores ingresos conforman la mayor proporción en el perfil de uso bajo. Además, se encuentra que en el perfil alto la proporción de hombres duplica a la de mujeres. En el perfil medio las mujeres son mayoría y en el perfil bajo son minoría, esto revela que las mujeres son las que más utilizan o aprovechan las herramientas de Internet.

Sábada (2010) realizó un estudio sobre el perfil de usuario de Internet en España, en el mismo recurre a varias fuentes de información para definir los perfiles de usuarios. Se plantean cuatro grupos. En el grupo de servicios de comunicación se incluyen usos como: correo electrónico, telefonía o video llamadas, chats, *blogs* o mensajería. Otro grupo de servicios de búsqueda de información y servicios. Un tercero de banca y compra-venta de servicios y el último grupo corresponde a usos de educación y formación. Luego de analizar los perfiles se concluye que para el caso de España existe una brecha digital marcada por la edad y la clase social, personas mayores de 65 años y personas de clase social baja no tienen representatividad dentro de los perfiles de usuarios de Internet.

Torres-Díaz et al. (2016) investigaron la incidencia del uso de Internet sobre el éxito académico. En primer lugar; se hace una clasificación de los estudiantes según el uso de Internet utilizando un análisis factorial exploratorio. Se obtienen dos grandes componentes, uno de uso de Internet en actividades académicas que incluye tres factores: comunicación, participación y búsqueda.

El otro componente lo nombraron uso de Internet para entretenimiento. Sin embargo; en este componente se tiene todo lo no relacionado con la parte académica. Dentro de este último se tienen tres factores, un factor llamado socialización, que incluye variables como postear en redes sociales, comentar, chatear y subir fotos o videos. Otro factor de descargas ya que se relaciona con variables de descarga de programas, música, películas, ver televisión y escuchar radio.

El tercer factor de transacciones y juegos incluye variables de compra y venta, así como de juegos en línea. Tomando como base los seis factores y utilizando un análisis de clúster se crean tres perfiles de usuarios de Internet. Se concluye que el uso de Internet en actividades académicas presenta los mismos patrones entre hombres y mujeres. Además, las mujeres utilizan en menor proporción el Internet para entretenimiento, dentro del componente de entretenimiento, las mujeres prefieren la descarga de información en lugar de los juegos en línea.

Agudo et al. (2012) realizaron un estudio en personas mayores, uno de los objetivos es describir los tipos de usos que hacen de las TIC, utilizando criterio de expertos se agrupan estos usos en cuatro categorías: formación, información, comunicación y entretenimiento. Se concluye que dentro de la categoría de formación y entretenimiento no existen diferencias entre hombres y mujeres, edad, estado civil, nivel de estudios, ni lugar de residencia. En las categorías de información y comunicación las mujeres sin estudios son las que menos hacen uso, las personas con estudios secundarios y universitarios presentan porcentajes más altos en el uso de estos recursos.

Cristóbal-Fransi et al. (2014) desarrollaron un estudio de segmentación de los usuarios de Internet, el objetivo fue analizar el comportamiento del internauta según los principales usos que piensa realizar de las diversas aplicaciones de Internet. Realizan un análisis factorial para clasificar a los usos en tres factores, el primero; nombrado *practicidad* tiene actividades relacionadas con gestiones bancarias, declaración de renta, solicitud de documentos a la administración pública, compras en línea, preparar viaje y realizar cursos.

El segundo factor llamado *socialización*, se refiere a participar en redes sociales, juegos en línea, buscar relaciones de amistad o pareja y vender cosas. Por último, en el factor de información, se tienen variables como: ver televisión, buscar trabajo, visitar sitios de museos, buscar información médica. Se concluye que el primer factor está formado por gente joven, con estudios y cierto nivel de ingresos. El segundo compuesto principalmente por género

femenino, edad madura y con nivel de estudios más básico. El tercero conformado por individuos de mediana edad, con estudios superiores e ingresos por encima de la media.

Los estudios realizados en esta línea brindan evidencia teórica para realizar la construcción de parcelas. Además, incluyen el análisis factorial exploratorio (AFE) como técnica de reducción de datos, y utilizan el análisis de conglomerados para crear de perfiles.

IV. METODOLOGÍA

En este apartado se describe la fuente de los datos, las variables que se utilizarán en el análisis, la principal técnica estadística empleada, así como los paquetes estadísticos utilizados para el análisis.

4.1 Fuente de datos

Los datos utilizados en el estudio corresponden la “Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017”. El trabajo de campo de la encuesta fue realizado por la Unidad de Servicios Estadísticos (USES) de la Escuela de Estadística de la Universidad de Costa Rica (UCR).

La encuesta tiene cobertura nacional y se definen como dominios de estudio las seis regiones de planificación: Central, Chorotega, Pacífico Central, Brunca, Huetar Caribe y Huetar Norte. La población de estudio son todas las personas residentes en Costa Rica, de 18 años o más, que residen habitualmente en viviendas particulares a noviembre de 2017. La unidad informante definida en la encuesta corresponde a toda persona residente en el país de 18 años o más. Para la selección del informante se utilizó el criterio de la persona más próxima en cumplir años (MICITT, 2019).

El diseño muestral es estratificado, bietápico, con selección PPT (probabilidad proporcional al tamaño), utilizando como medida de tamaño el total de viviendas de cada UPM. La muestra

planeada fue de 3 500 entrevistas, y se asignaron 1000 a la Región Central y 500 en cada una de las cinco regiones restantes. Para la selección de la muestra se utilizó el Marco Muestral de Viviendas 2011 (MMV-2011) del Instituto Nacional de Estadística y Censos (INEC).

4.2 Variables de estudio

En el siguiente cuadro se tiene la distribución de las variables sociodemográficas, cabe destacar que a la base de datos se le aplica una corrección por no respuesta y se toma como base la ENAHO 2017 para realizar el ajuste por sexo y grupos de edad.

Cuadro 1. Distribución porcentual de las variables sociodemográficas de la EAU 2017

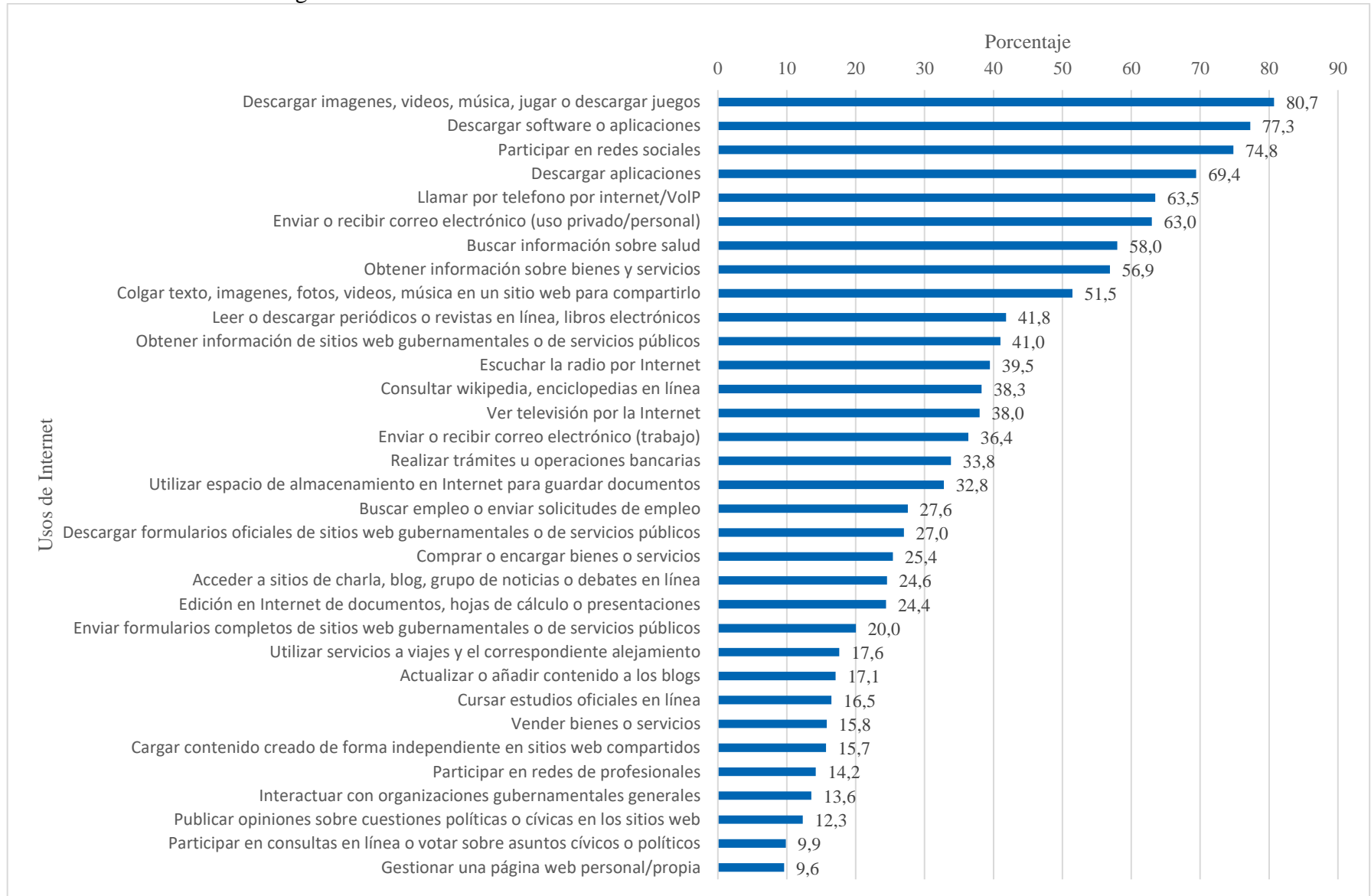
Variables	Frecuencia	%
Total	3 618	100
Región		
Central	1 006	27,8
Chorotega	492	13,6
Pacífico Central	520	14,4
Brunca	495	13,7
Huetar Atlántico	508	14,0
Huetar Norte	598	16,5
Sexo		
Masculino	1 729	47,8
Femenino	1 889	52,2
Grupos de edad		
18 a 24	586	16,2
25 a 34	755	20,9
35 a 44	641	17,7
45 a 54	610	16,9
55 o más	1 026	28,3
Educación		
Primaria y menos	1 644	45,4
Secundaria	1 331	36,8
Universitaria	609	16,8

NS/NR	6	0,2
Ingresos		
Menos de 250 mil colones	1 438	39,8
Entre 250 mil y menos de 500 mil	1 229	34,0
Entre 500 mil y menos de 750 mil	363	10,0
Más de 750 mil	409	11,3
NS/NR	179	4,9
Ingresos subjetivos		
No les alcanza, tienen grandes dificultades	340	9,4
No les alcanza, tiene dificultades	1 197	33,1
Les alcanza justo, sin grandes dificultades	1 520	42,0
Les alcanza bien, pueden ahorrar	530	14,7
NS/NR	31	0,9

Fuente: MICITT. Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017.

En el Gráfico 1 se presenta la distribución porcentual de las 33 variables que se utilizarán en el análisis, los ítems están relacionados con diferentes usos que se le pueden dar a Internet, si utiliza o no Internet para cada una de las actividades descritas, el porcentaje de no respuesta en todos los ítems es de 0,3 %, al ser un porcentaje tan bajo no se realiza ningún tratamiento especial a los datos y se consideran todas las variables en el análisis.

En el gráfico se observa como la proporción de personas que realizan las diferentes actividades es variable, las actividades de más incidencia se presentan en 80 % de las personas y las de menos frecuencia llegan a observarse en aproximadamente 10 % de las personas entrevistadas.

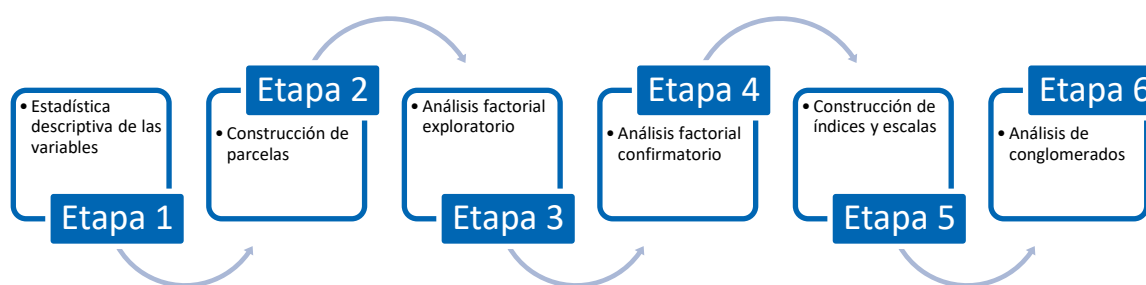
Gráfico 1. Uso de Internet según actividad

Fuente: MICITT (2017). Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017. [Archivo de datos]

4.3 Técnicas estadísticas

En la Figura 1 se detallan las etapas del análisis y las técnicas estadísticas empleadas. En primera instancia se realizó un análisis descriptivo de las variables, se procedió con la construcción de parcelas, se realizó un análisis factorial exploratorio, seguidamente se realizó la construcción del índice y las escalas. En la etapa cinco se formuló el análisis factorial confirmatorio, y por último se realizó un análisis de conglomerados.

Figura 1. Etapas empleadas en el análisis y técnicas estadísticas utilizadas



Luego de realizar el análisis descriptivo de las 33 variables, como segundo paso se realiza una categorización en parcelas, la idea de esta técnica es agrupar variables de una temática similar. Nunnally (1994) expone que una de las herramientas psicométricas para mejorar la precisión en los resultados es la construcción de parcelas, esto siempre y cuando los ítems presenten la misma calidad técnica y correspondan a escalas.

Para probar la consistencia interna de los ítems se utiliza un análisis de fiabilidad en cada una de las parcelas por medio del coeficiente alfa de Cronbach, así como una revisión de las cargas factoriales de cada uno de los ítems utilizando un AFE en cada parcela.

El alfa de Cronbach mide la confiabilidad de la consistencia interna, es decir; el grado en que las respuestas son consistentes entre los elementos de una medida (Kline, 2011). Según Nunnally (1994) para que al valor del alfa se considere aceptable debe ser mayor a 0,7, por debajo de ese valor la consistencia interna de la escala es baja. Sin embargo; según Oviedo et al. (2005) se pueden aceptar valores inferiores de alfa de Cronbach teniendo presente esa limitación.

El AFE consiste en una técnica multivariable empleada para reducir datos, el objetivo es, representar en un número más reducido de variables hipotéticas, un conjunto de variables cuantitativas observadas (Finch, 2020). Se utiliza una rotación oblicua ya que según Méndez (2012) se parte del supuesto de que los factores están asociados y en la práctica es el escenario más común. Además, al tratarse de variables dicotómicas se utiliza la matriz de correlaciones tetracórica como un método más robusto para la estimación de los resultados (Martínez-Abad et al., 2017).

En la cuarta etapa se realiza la construcción del índice y las escalas, según Gómez (citado en Montero, 2008:16) los índice y escalas son medidas compuestas, el índice relaciona variables o constructos de diferente naturaleza, mientras que una escala relaciona variables de la misma naturaleza. Según señala Montero (2008) en la escala el constructo está claramente definido, mientras que en el índice los constructos y variables podrían estar menos definidos desde un punto de vista teórico. La misma autora señala como en la construcción de escalas es necesario el uso de métodos psicométricos para evidenciar la validez y confiabilidad, se dice que una de las medidas más importantes para establecer la confiabilidad es el alfa de Cronbach.

La quinta etapa corresponde a la construcción del análisis factorial confirmatorio (AFC), actualmente, es uno de los procedimientos estadísticos más utilizados en la investigación, Orgaz (2008) indica que el objetivo de este es determinar por medio de pruebas cuantitativas, en qué medida los datos de la muestra apoyan un modelo teórico de múltiples relaciones de dependencia entre variables.

El AFC es un caso particular de los modelos de ecuaciones estructurales (SEM, por sus siglas en inglés). Según Martínez (citado en Fernández 2008:32), este análisis consiste en un modelo de medida que intenta explicar la manera en que un conjunto de variables empíricas explica otras variables latentes (no observables empíricamente).

Tal y como lo exponen Brown (2006) y Fernández (2008), en este análisis el investigador debe tener a priori evidencia teórica o alguna hipótesis establecida con el fin de comprobarla por medio de este modelo de medida. De ahí que, previamente se plantea un número determinado de factores (variables latentes) y los indicadores o variables observados que se incluyen en cada uno de ellos. Fernández (2008) resalta además que en un AFC se puede plantear un modelo donde los factores tengan o no relación con otros factores. Además, las correlaciones entre los errores se pueden considerar como parte del modelo.

Según lo plantean Fernández (2008), una de las formas de denotar los modelos es por medio de un conjunto de ecuaciones lineales, el cual se puede definir como:

$$X = \Lambda_x \xi + \delta.$$

Donde:

X = vector $p \times 1$ de puntuaciones diferenciales en las variables observadas

Λ_x = matriz $p \times$

k de las saturaciones de las p variables en los k factores comunes

ξ = vector $k \times 1$ de puntuaciones diferenciales de factores comunes

δ = vector $p \times 1$ de factores únicos y error de cada una de las variables observadas

Otra manera de representar cualquier modelo según lo señala Fernández (2008) así como Kline (2011), Brown (2006) y Mulaik (2009); es por medio del *Path Diagram* o Diagrama de Relaciones, y se utiliza la siguiente simbología:

- Variables latentes o factores en círculos u óvalos.
- Variables observadas en cuadrados o rectángulos.

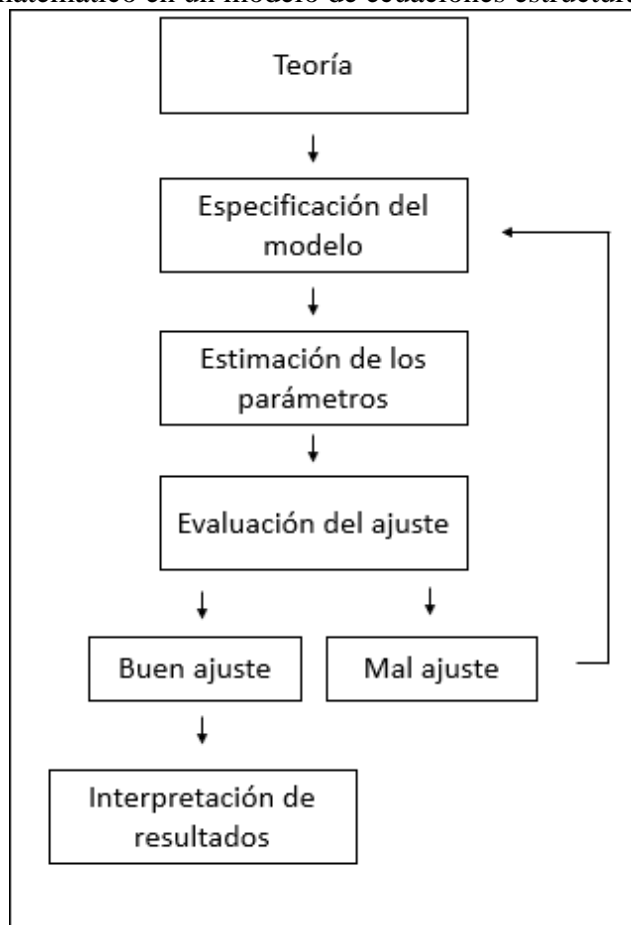
- Relaciones de dependencia por medio de flechas unidireccionales que van de los factores a las variables observadas.
- Correlaciones o covariancias por medio de flechas bidireccionales.

Tal y como lo indica Orgaz (2008) para realizar un modelo de ecuaciones estructurales es fundamental el planeamiento de un modelo matemático y la teoría en la que se basa. El modelo matemático consta de tres fases; la especificación del modelo, la estimación de los parámetros y la evaluación del ajuste del modelo.

Tomando como punto de partida la teoría y los datos, se realiza la especificación del modelo en la cual se determinan las variables, la relación entre ellas y la dirección y magnitud de las relaciones; según recalca Kline (2011) este es el paso más importante, ya que, las siguientes etapas suponen que esto es correcto, se sugiere realizar un mapeo de posibles cambios en el modelo, sustentado siempre por la teoría, ya que, en caso de que el modelo no ajuste se debe volver a su especificación.

El siguiente paso es la estimación de los parámetros, para esto se necesita el uso de una herramienta informática para realizar el análisis. Seguidamente, la evaluación del ajuste donde se analiza, que tan bien el modelo teórico propuesto puede ser explicado con los datos que se tienen, de presentar un buen ajuste se procede a la interpretación de los resultados, caso contrario tal y como se había mencionado, se debe volver a la especificación del modelo. En la Figura 2 se muestra en resumen este proceso por seguir en el planteamiento de un modelo de ecuaciones estructurales. En el Anexo 1 se muestra un diagrama de flujo de los pasos para un modelo SEM según lo expone Kline (2011).

Figura 2. Proceso para realizar el planteamiento matemático en un modelo de ecuaciones estructurales



Fuente: Adaptación de Orgaz. (2008). Introducción a la Metodología SEM: Concepto y Propósitos Fundamentales (p.22).

En la evaluación del ajuste de un modelo de ecuaciones estructurales, se busca ajustar las covarianzas entre las variables, es decir; minimizar la diferencia entre las covarianzas observadas en la muestra y las covarianzas pronosticadas por el modelo estructural, Ruiz et.al. (2010). Para esto se utilizan estadísticos de bondad de ajuste, se tienen tres tipos, de ajuste absoluto (valoran los residuos o diferencias entre las matrices observadas y estimadas), de ajuste incremental o comparativo (comparan el ajuste del modelo propuesto con otro de ausencia de relación entre las variables o modelo de independencia) y de ajuste parsimonioso (valoran el ajuste con respecto al número de parámetros estimados). Estos estadísticos han sido propuestos por Kaplan (2009), Kline (2011), Ruiz et.al. (2010), Brown (2006), Mulaik (2009), Orgaz (2008), entre otros.

Cuadro 2. Estadísticos de bondad de ajuste y criterios en la evaluación SEM

	Estadístico	Abreviatura	Criterio
Ajuste absoluto	Chi-cuadrado 1/	χ^2	$p < 0,05$
	Residuo Cuadrático Medio (Root Mean Square Residual)	RMR o RMSR	Próximo a cero
	Error de Aproximación (Root Mean Square Error of Aproximation)	RMSEA	$\leq 0,05$
	Criterio de Información de Akaike (Akaike Information Criterion)	AIC	Valores bajos
Ajuste comparativo	Índice Ajustado de Bondad de Ajuste (Adjusted Goodness of Fit Index)	AGFI	$\geq 0,90$
	Índice de Ajuste Comparativo (Comparative Fit Index)	CFI	$\geq 0,90$
	Índice de Ajuste Normalizado (Bentler-Bonnet Normed Fit Index)	NFI	$\geq 0,90$
	Índice de Ajuste No Normalizado (BentlerBonnet Non-Normed Fit Index)	NNFI	$\geq 0,90$
Ajuste de parsimonia	Residuo Cuadrático Medio Estandarizado (Standardized Root Mean Squared Residual)	SRMR	$\leq 0,08$
	Índice de Bondad de Ajuste (Goodness of Fit Index)	GFI	$\geq 0,90$
	Índice de Ajuste Parsimonioso (Parsimonius Fit Index)	PFI	Próximo a 1
	Coefficiente de Determinación (The Coefficient of Determination)	CD	$->1$

1/ Prueba muy sensible al tamaño de muestra

Fuente: Adaptación de Kline (2011), Ruiz et.al. (2010), Brown (2006), Mulaik (2009), Orgaz (2008), Kaplan (2009).

En la última etapa, se realiza un análisis de conglomerados utilizando las escalas para así crear perfiles de usuarios e interpretar de una manera más resumida los datos. Según recalca Yang (2010) este análisis es una herramienta exploratoria donde se intenta dividir los casos en pocos grupos con respecto a un conjunto de variables, así como el significado de estos grupos, se dice que la principal razón para el uso de este análisis es encontrar grupos de entidades similares en una muestra de datos (Aldenderfer Mark & Blashfield Roger, 1984). No existe un conjunto correcto o incorrecto, corresponde a una interpretación que por muchas

razones y según los casos en el conjunto de datos pueden ser mejores o peores (Uprichard Emma, 2009).

Este mismo autor señala que el análisis de conglomerados es una herramienta muy poderosa, tiene la capacidad de reducir cientos, miles y millones de casos a un conjunto manejable, menciona que la clasificación funciona si se encuentran diferencias significativas y estas son de utilidad y representan alguna característica de la organización de los datos.

4.4 Paquetes estadísticos

Para el análisis y procesamiento de los datos se utilizaron los programas estadísticos de análisis de datos, *STATA* (Statistical Software for Data Science), *SPSS* (*Statistical Package for the Social Sciences*) y *Rstudio*.

V. RESULTADOS

Tal y como se explicó, el análisis de la información se realizó en varias etapas, en primer lugar; se construyeron parcelas a partir de los ítems, se realizó un análisis factorial exploratorio, así como la construcción del índice y las escalas, seguidamente se realizó un análisis factorial confirmatorio y por último un análisis de conglomerados con el fin de realizar una caracterización más específica.

5.1 Construcción de parcelas y análisis factorial exploratorio

Tomando como base el sustento teórico, se realiza la agrupación de los ítems en parcelas. Se encontraron cinco áreas de contenido dentro de los usos de Internet, estas son: Informativo, Interacción social, Entretenimiento, Productivo y Habilitación Ciudadana.

En cada agrupación se verifica evidencia de unidimensionalidad; para esto se examinó el coeficiente alfa de Cronbach y las cargas factoriales de cada uno de los ítems.

Obsérvese en el Cuadro 3 como a lo interno de cada una de las parcelas todos los ítems muestran un peso relativo por encima de 0,3, además en el análisis de fiabilidad 4 de las 5 parcelas muestra valores favorables superiores o iguales a 0,7, únicamente el factor de Interacción social presenta un alfa 0,6. Sin embargo; tomando como base el análisis teórico se considera aceptable para esta investigación, siempre teniendo presente que la consistencia interna de esa escala es un poco baja. Como parte de los resultados que da este análisis es la evaluación individual de cada uno de los ítems y si este contribuye o no en la confiabilidad de la parcela, en este caso ninguno de los ítems presentaba una condición desfavorable por lo que no fue necesario eliminar ninguna de las variables de estudio.

Cuadro 3. Cargas factoriales y análisis de fiabilidad según parcelas

Parcela	Ítem	Cargas factoriales	Alfa de Cronbach
Informativa	Obtener información sobre bienes y servicios	0,93	0,7
	Buscar información sobre salud	0,93	
Interacción social	Llamar por teléfono por internet/VoIP	0,62	0,6
	Participar en redes sociales	0,73	
	Acceder a sitios de charla, blog, grupo de noticias o debates en línea	0,73	
	Cargar contenido creado de forma independiente en sitios web compartidos	0,66	
	Colgar texto, imágenes, fotos, videos, música en un sitio web para compartirlo	0,78	
	Actualizar o añadir contenido a los blogs	0,80	
	Entretenimiento	Escuchar la radio por internet	
Ver televisión por la internet		0,77	
Descargar imágenes, videos, música, jugar o descargar juegos		0,77	
Descargar aplicaciones		0,83	
Descargar software o aplicaciones		0,86	
Leer o descargar periódicos o revistas en línea, libros electrónicos		0,76	
Productividad	Enviar o recibir correo electrónico (trabajo)	0,80	0,9
	Enviar o recibir correo electrónico (uso privado/personal)	0,88	
	Comprar o encargar bienes o servicios	0,79	
	Vender bienes o servicios	0,70	
	Utilizar servicios a viajes y el correspondiente alejamiento	0,78	
	Realizar trámites u operaciones bancarias	0,82	

	Cursar estudios oficiales en línea	0,80	
	Consultar Wikipedia, enciclopedias en línea	0,71	
	Buscar empleo o enviar solicitudes de empleo	0,62	
	Participar en redes de profesionales	0,80	
	Gestionar una página web personal/propia	0,65	
	Espacio de almacenamiento en Internet para guardar documentos	0,79	
	Edición en Internet de documentos, hojas de cálculo o presentaciones	0,82	
	Obtener información de sitios web gubernamentales o de servicios públicos	0,92	
	Descargar formularios oficiales de sitios web gubernamentales o de servicios públicos	0,94	
Habilitación ciudadana	Enviar formularios completos de sitios web gubernamentales o de servicios públicos	0,92	0,8
	Interactuar con organizaciones gubernamentales generales	0,92	
	Publicar opiniones sobre cuestiones políticas o cívicas en los sitios web	0,73	
	Participar en consultas en línea o votar sobre asuntos cívicos o políticos	0,78	

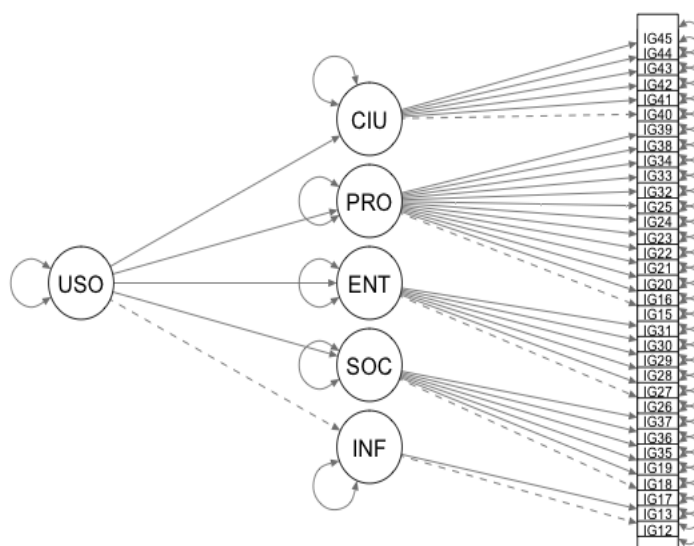
Fuente: MICITT (2017). Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017. [Archivo de datos].

5.2 Análisis factorial confirmatorio

Para la elaboración del modelo confirmatorio se trabajó con los indicadores construidos a partir de las parcelas. Al tratarse de variables dicotómicas se debe estimar el modelo utilizando la matriz de correlaciones tetracórica. Además, por la naturaleza de los datos, se podría violentar el supuesto de normalidad multivariante para una estimación máximo verosímil, por lo tanto, para realizar las estimaciones de los coeficientes se utiliza el método de estimación por mínimos cuadrados ponderados (DWSL por sus siglas en inglés: Diagonally Weighted Least Squares).

El planteamiento del modelo se detalla en la Figura 3. Obsérvese que se tiene el constructo de uso de Internet el cual se va a explicar con los 5 indicadores construidos.

Figura 3. Planteamiento del modelo a estimar en el AFC

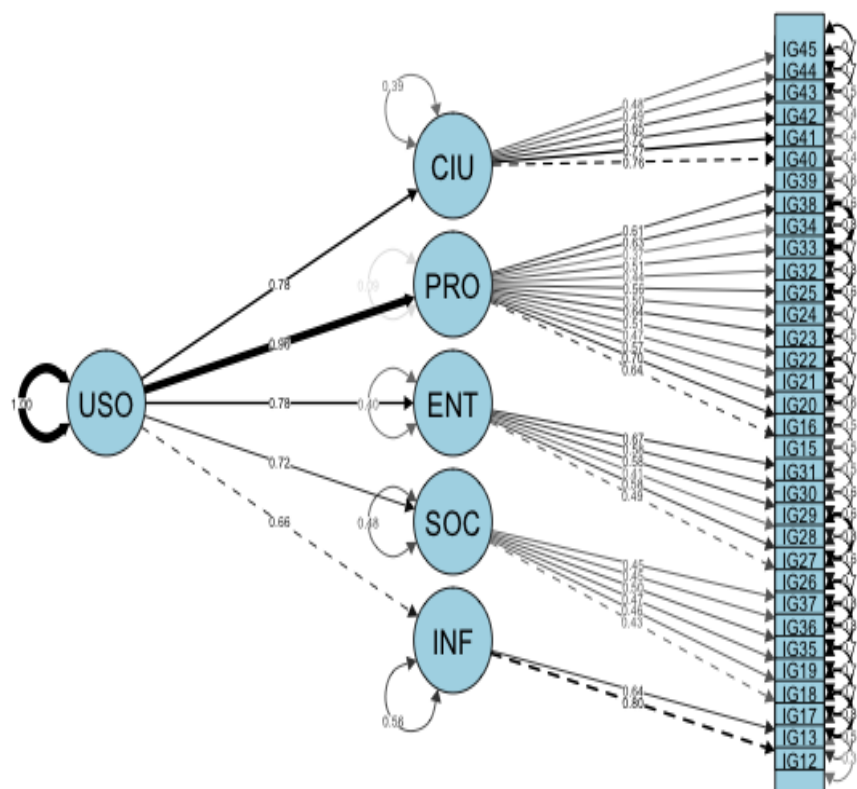


Nota: USO=Uso de Internet, CIU=Habilitación ciudadana, PRO=Productividad, ENT=Entretenimiento, SOC=Interacción social, INF=Informativa.

Fuente: MICITT (2017). Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017. [Archivo de datos].

En la Figura 4 se tiene el diagrama de relaciones que resume los resultados de las estimaciones estandarizadas tanto del constructo con cada uno de los factores o parcelas, así como la relación de estos con cada una de las variables correspondientes, en el Cuadro 4 y Cuadro 5 se presentan los resultados de este diagrama.

Figura 4. Estimación del modelo de uso de Internet por medio del diagrama de relaciones



Nota: USO=Uso de Internet, CIU=Habilitación ciudadana, PRO=Productividad, ENT=Entretenimiento, SOC=Interacción social, INF=Informativa.

Fuente: MICITT (2017). Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017. [Archivo de datos].

En el Cuadro 4 se observa que la relación entre las parcelas y las variables tienen cargas factoriales superiores a 0,3, esto indica que las variables tienen una relación significativa con las parcelas o factores correspondientes. Además, en el Cuadro 5 se aprecia la relación entre las parcelas y la variable no observada, las cuales presentan igualmente cargas factoriales superiores a 0,3.

Cuadro 4. Estimación del modelo de uso de Internet, relación entre parcelas y variables

Parcela	Ítem	Resultados del modelo		
		Coefficientes estándar	Error estándar	Z
Informativa	Obtener información sobre bienes y servicios	0,797	0,015	51,475
	Buscar información sobre salud	0,639	0,012	51,175
Interacción social	Llamar por teléfono por internet/VoIP	0,429	0,010	44,686
	Participar en redes sociales	0,460	0,010	46,550
	Acceder a sitios de charla, blog, grupo de noticias o debates en línea	0,475	0,012	39,502
	Cargar contenido creado de forma independiente en sitios web compartidos	0,504	0,016	32,075
	Colgar texto, imágenes, fotos, videos, música en un sitio web para compartirlo	0,449	0,010	46,150
	Actualizar o añadir contenido a los blogs	0,451	0,014	31,692
Entretenimiento	Escuchar la radio por internet	0,492	0,010	51,516
	Ver televisión por la internet	0,579	0,010	55,641
	Descargar imágenes, videos, música, jugar o descargar juegos	0,411	0,009	45,139
	Descargar aplicaciones	0,579	0,009	65,257
	Descargar software o aplicaciones	0,576	0,009	61,076
	Leer o descargar periódicos o revistas en línea, libros electrónicos	0,666	0,010	65,139
Productividad	Enviar o recibir correo electrónico (trabajo)	0,642	0,009	68,086
	Enviar o recibir correo electrónico (uso privado/personal)	0,699	0,007	99,209
	Comprar o encargar bienes o servicios	0,568	0,012	47,668
	Vender bienes o servicios	0,465	0,012	37,400
	Utilizar servicios a viajes y el correspondiente alejamiento	0,514	0,014	37,808
	Realizar trámites u operaciones bancarias	0,639	0,010	61,083
	Cursar estudios oficiales en línea	0,499	0,015	34,348
	Consultar Wikipedia, enciclopedias en línea	0,559	0,009	59,491
	Buscar empleo o enviar solicitudes de empleo	0,445	0,011	41,867
	Participar en redes de profesionales	0,513	0,015	33,527
Habilitación ciudadana	Gestionar una página web personal/propia	0,369	0,015	25,007
	Espacio de almacenamiento en Internet para guardar documentos	0,633	0,011	56,530
	Edición en Internet de documentos, hojas de cálculo o presentaciones	0,606	0,015	41,728
Habilitación ciudadana	Obtener información de sitios web gubernamentales o de servicios públicos	0,764	0,010	73,489
	Descargar formularios oficiales de sitios web gubernamentales o de servicios públicos	0,773	0,015	51,824

Enviar formularios completos de sitios web gubernamentales o de servicios públicos	0,717	0,016	43,999
Interactuar con organizaciones gubernamentales generales	0,648	0,018	35,974
Publicar opiniones sobre cuestiones políticas o cívicas en los sitios web	0,486	0,016	31,245
Participar en consultas en línea o votar sobre asuntos cívicos o políticos	0,481	0,018	26,836

Fuente: MICITT (2017). Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017. [Archivo de datos].

Cuadro 5. Estimación del modelo de uso de Internet, relación entre el constructo y las parcelas

Constructo	Parcela	Resultados del modelo		
		Coefficientes estándar	Error estándar	Z
Uso de Internet	Informativa	0,665	0,013	50,836
	Interacción social	0,722	0,011	66,762
	Entretenimiento	0,776	0,008	91,766
	Productividad	0,955	0,008	12,493
	Habilitación ciudadana	0,783	0,009	91,869

Fuente: MICITT (2017). Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017. [Archivo de datos].

Para evaluar el modelo se analizan los estadísticos de bondad de ajuste utilizando los criterios de la evaluación SEM, los resultados se presentan en el Cuadro 6, se aprecia que el modelo final cumple con los criterios, por lo tanto, se comprueba que presenta un buen ajuste.

Cuadro 6. Estadísticos de ajuste del AFC

Estadístico	Criterio	Resultado del AFC
Chi-cuadrado	$p < 0,05$	0,000
CFI	$\geq 0,90$	0,988
NNFI	$\geq 0,90$	0,987
RMSEA	$\leq 0,05$	0,046
SRMR	$\leq 0,08$	0,072

Fuente: MICITT (2017). Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017. [Archivo de datos].

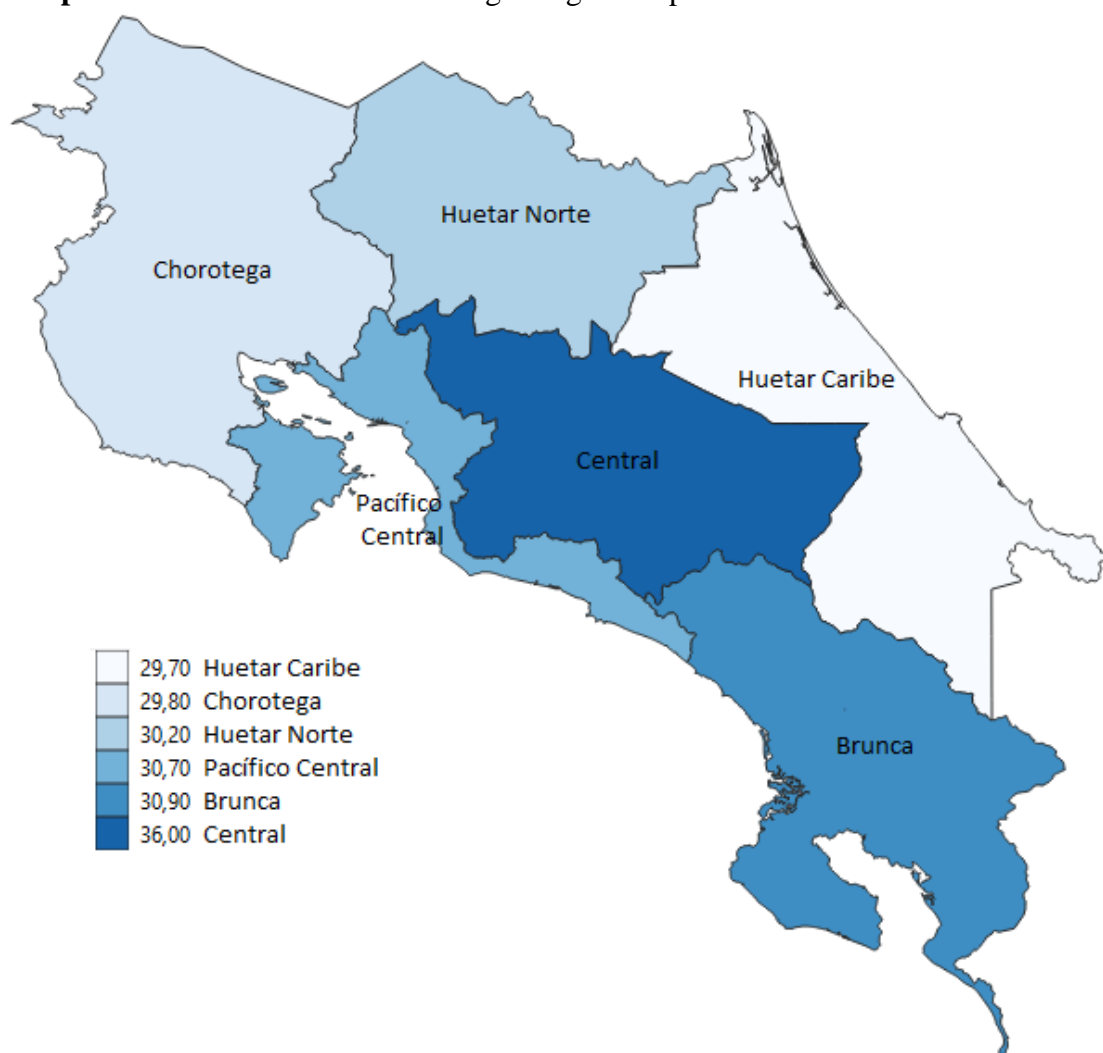
5.3 Construcción del índice y escalas

Una vez validado el modelo y a partir de las parcelas identificadas, se crean las siguientes escalas: Informativa, Interacción social, Entretenimiento, Productividad y Habilitación Ciudadana. El cálculo se realiza sumando cada una de las variables que la componen y el resultado se estandariza de manera que oscile entre 0 y 100. Para el cálculo del índice de uso de Internet se procede de la misma manera, todos los factores tienen el mismo peso por lo que se realiza la suma de las variables y se estandarizan los resultados.

En el Cuadro 7 se observa la media y los intervalos de confianza del índice de uso de Internet, el cálculo se realiza tomando el diseño muestral complejo. Cabe recalcar que el índice oscila entre 0 y 100 y presenta una media nacional de 31,9.

Para la región Central la media del índice es mayor al límite superior de la media nacional, por lo tanto, en esta región el índice de uso de Internet es significativamente mayor. En las regiones: Chorotega, Pacífico Central, Huetar Caribe y Huetar Norte, sucede lo contrario, la media del índice está por debajo del límite inferior de la media nacional, lo que refleja una diferencia significativamente menor en estas regiones. La región Brunca es la única que no presenta diferencias significativas. Estos resultados muestran una clara diferenciación entre la región Central y el resto de las regiones con respecto al índice de uso de Internet, es decir; en la región Central se tiene una mayor incidencia en los usos que se le dan a Internet (ver mapa 1)

Mapa 1. Índice de uso de Internet según región de planificación



Fuente: MICITT (2017). Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017. [Archivo de datos].

Cuando se observa la media del índice por sexo, los hombres presentan una diferencia significativamente mayor y por su parte, las mujeres tienen una media por debajo del límite inferior del promedio nacional, resultados que dejan clara la brecha existente en los usos de Internet entre hombres y mujeres.

Por grupos de edad también se aprecian diferencias significativas, las personas entre los 18 y 44 años de edad tienen un promedio en el índice de uso de Internet significativamente mayor, mientras que las personas de más de 45 años, presentan una media significativamente

inferior, estas diferencias demuestran la brecha generacional existente con respecto a los usos de Internet.

Con respecto al nivel educativo, se aprecia que las personas con primaria o menos tienen una media significativamente inferior; aquellos que tienen nivel educativo de secundaria no presentan diferencias con respecto a la media nacional y las personas con nivel educativo universitario poseen una media en el índice significativamente superior.

Analizando los resultados del índice por nivel de ingresos, se aprecia que las personas cuyo ingreso se encuentra por debajo de los 250 mil colones presentan una media del índice de uso de Internet de 22,1 inferior al promedio nacional 31,9, esta diferencia es significativa. Mientras que las personas cuyos ingresos superan los 500 mil colones cuentan con una media del índice por encima del límite superior en el nivel nacional. Estos resultados se refuerzan al analizar los datos utilizando el nivel de ingresos subjetivos, donde las personas que dicen no les alcanza presentan una diferencia significativamente menor, mientras que aquellos que indican les alcanza justo o bien y pueden ahorrar tienen una diferencia significativamente mayor.

En este sentido se puede concluir que el valor del índice de uso de Internet depende de la zona de residencia (región de planificación), el sexo, la edad, el nivel de educación y el nivel de ingresos de las personas; en todas las variables sociodemográficas evaluadas existen brechas marcadas.

Cuadro 7. Media e intervalos de confianza del índice de uso de Internet según variables sociodemográficas

Media nacional		31,9
Intervalos de confianza	Límite superior	33,1
	Límite inferior	30,8
Región de planificación	Central	36,0
	Chorotega	29,8
	Pacífico Central	30,7
	Brunca	30,9
	Huetar Caribe	29,7
	Huetar Norte	30,2
Sexo	Hombre	35,6

	Mujer	28,6
Grupos de edad	18 a 24	37,3
	25 a 34	38,5
	35 a 44	33,2
	45 a 54	25,6
	55 o más	17,9
Educación	Primaria o menos	19,9
	Secundaria	31,7
	Universitaria	50,1
Ingreso	Menos de 250 mil colones por mes	22,1
	Entre 250 mil y menos de 500 mil	31,1
	Entre 500 mil y menos de 750 mil	38,1
	Más de 750 mil	49,6
Ingreso subjetivo	No les alcanza, tienen grandes dificultades	22,2
	No les alcanza, tiene dificultades	24,1
	Les alcanza justo, sin grandes dificultades	33,4
	Les alcanza bien, pueden ahorrar	44,1

Nota: Se presentan diferencias significativas en todas las variables sociodemográficas.

Fuente: MICITT (2017). Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017. [Archivo de datos].

El índice de uso de Internet está compuesto por cinco escalas: Informativa, Interacción Social, Entretenimiento, Productividad, y Habilidad ciudadana. En el Cuadro 8 se observa la media y los intervalos de confianza para cada una de las escalas construidas. Cabe destacar que las escalas que presentan una media más alta corresponden a entretenimiento, la escala informativa y la de interacción social.

Cuadro 8. Media e intervalos de confianza de las escalas de uso de Internet

Escalas	Media	Intervalo de confianza	
		95%	
		Inferior	Superior
Informativa	52,8	50,4	55,1
Interacción social	39,3	38,3	40,4
Entretenimiento	52,4	51,0	53,7
Productividad	22,0	20,6	23,5
Habilidad ciudadana	18,5	17,2	19,8

Fuente: MICITT (2017). Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017. [Archivo de datos].

En el Anexo 2 se resumen los resultados de la escala informativa con respecto a las variables sociodemográficas, esta escala está compuesta principalmente de las actividades de obtener información sobre bienes y servicios, así como de obtener información sobre salud.

En la región Central existe una diferencia significativamente mayor con respecto al promedio nacional, mientras que, para las regiones Chorotega, Pacífico Central y Huetar Norte los resultados se encuentran por debajo del promedio nacional, en el caso de la región Brunca y Huetar Caribe no presentan diferencias ya que, se ubica dentro de los valores promedio de la escala nacional.

Esta escala no presenta diferencias significativas entre hombres y mujeres. Por grupos de edad llama la atención que las personas entre los 25 y 44 años tienen un promedio superior, hacen un mayor uso de estas actividades, mientras que las personas de más de 45 años presentan valores significativamente por debajo de la media, siendo el grupo de edad de 55 años o más los que hacen un menor uso de estas actividades.

Con respecto al nivel educativo las diferencias son marcadas, las personas con educación universitaria presentan valores muy por encima del promedio nacional, mientras que las personas sin educación o con nivel de instrucción de primaria poseen valores muy bajos, comportamiento que se repite en cuanto al nivel de ingresos, ya que las personas con ingresos superiores a los 500 mil colones poseen valores significativamente mayores y aquellos con ingresos de 250 mil o menos presentan valores por debajo de la media nacional.

El Anexo 3 resume los resultados de la escala de interacción social, esta incluye variables relacionadas con la participación en redes sociales, blogs, cargar contenido, texto, imágenes, fotos, videos en sitios web, así como llamadas en línea.

La región Chorotega es la que presenta una media más alta en esta escala, superando significativamente a la media nacional junto con las regiones Pacífico Central y Brunca, las regiones Central, Huetar Caribe y Huetar Norte tiene un valor significativamente más bajo. Entre hombres y mujeres no se presentan diferencias significativas.

La brecha generacional también está presente en esta escala, siendo las personas más jóvenes (18 a 34 años) las que presentan valores significativamente más altos y los de mayor edad (más de 45 años) una media inferior a la nacional. Con respecto al nivel de ingresos, se repite la tendencia, las personas con ingresos inferiores a los 250 mil colones presentan valores en la escala de interacción social inferior, mientras que aquellos cuyos ingresos superan los 750 mil colones tienen un valor mayor en la escala.

En el Anexo 4 se presentan los resultados de la escala de entretenimiento, esta es una de las escalas que presenta la media nacional más alta, está compuesta por variables como escuchar radio, ver televisión, leer, y jugar en línea, descargar imágenes, videos, música, aplicaciones y software.

Las regiones Chorotega y Pacífico Central tienen una media significativamente superior a la media nacional, mientras que las regiones Huetar Caribe se encuentran por debajo. Las mujeres por su parte se encuentran por debajo del promedio nacional, y los hombres presentan valores superiores significativos. Se tiene además que las personas entre 18 y 34 años presentan valores superiores en esta escala, mientras que aquellas de más de 45 tienen valores inferiores por debajo de la media.

Aquellas personas sin ningún nivel de instrucción o con primaria tienen valores significativamente menores y quienes tienen secundaria o universitaria poseen valores por encima de la media nacional. El nivel de ingreso también discrimina en esta escala, quienes tienen ingresos inferiores a los 250 mil colones presentan una media inferior y aquellos que superan los 250 mil colones poseen una diferencia significativamente mayor.

Otra de las escalas que componen el índice de uso de Internet es la escala de productividad, en esta se agrupan todas las variables relacionadas con enviar y recibir correos, compra y venta de bienes y servicios, operaciones bancarias, llevar cursos, búsqueda de empleo, gestión de una página web, espacio de almacenamiento y edición de documentos en línea, en el Anexo 5 se resumen los resultados de acuerdo con las variables sociodemográficas.

Al analizar los resultados por región de planificación se tiene una clara diferencia entre la región Central y el resto de las regiones, donde la media es significativamente superior y en las restantes regiones, a excepción de la región Huetar Norte la cual se encuentra dentro del promedio, se aprecia un promedio por debajo del límite inferior en el nivel nacional. Además, los hombres presentan diferencias significativamente superiores y las mujeres se encuentran por debajo de la media nacional, lo que muestra una clara brecha entre hombres y mujeres.

Si se observan los datos por grupos de edad, se tiene que las personas de 18 a 34 años tienen una media por encima del promedio nacional, mientras que aquellas de más de 45 años presentan valores significativamente inferiores. El nivel de ingreso también presenta diferencias significativas, las personas cuyos ingresos son inferiores a los 500 mil colones tienen valores por debajo de la media nacional y quienes superan esta cifra presentan valores significativamente superiores.

Por su parte, en el Anexo 6 se presentan los resultados de la escala de habilitación ciudadana, esta escala se compone de variables relacionadas con la búsqueda de información gubernamental e interacción, descarga y envío de formularios de sitios web gubernamentales o de servicios públicos, publicar opiniones sobre asuntos políticos o cívicos, así como la participación en consultas o votar sobre estos asuntos.

En la región Central se tiene una diferencia significativamente superior, mientras que, en las regiones Chorotega, Pacífico Central, Huetar Caribe y Huetar Norte se aprecian valores por debajo de la media nacional. Entre hombres y mujeres existe una clara diferencia, los hombres presentan valores significativos por encima del promedio y las mujeres valores por debajo, la participación masculina en este tipo de actividades es notablemente superior.

Por grupos de edad se presenta un comportamiento particular, las personas con edades entre los 25 y 34 años tienen valores significativamente más altos y aquellas de 45 años o más poseen una media inferior a la media nacional, siendo los grupos de edad de menor participación en este tipo de actividades.

Por nivel de instrucción también se aprecian diferencias marcadas, las personas con nivel educativo de secundaria o menos presentan valores muy por debajo de la media nacional, mientras que las personas con nivel de instrucción universitaria tienen una media mucho mayor al valor nacional, lo que significa que tienen mayor participación en los usos asignados a esta escala de habilitación ciudadana.

La variable ingresos también presenta diferencias significativas, las personas con ingresos inferiores a los 500 mil colones tienen valores significativamente más bajos y aquellas que superan esta cifra poseen una media mayor.

5.4 Análisis de conglomerados

Con el fin de caracterizar de una manera más específica a los usuarios de Internet se crearon conglomerados utilizando las escalas que componen el índice. Al tratarse de variables métricas se utiliza el procedimiento k-medias para crear los conglomerados (ver Anexo 7).

Se obtiene una clasificación de dos conglomerados y dentro del contexto en que se realiza la investigación se les asignan los siguientes nombres:

- Consumidor social: este conglomerado representa 52 % de las personas entrevistadas, se caracteriza por participar en redes sociales, acceder a blog, noticias, debates, cargar y descargar contenido, escuchar radio y ver televisión en línea.
- Ciudadano digital: el conglomerado abarca 48 % de los entrevistados, este grupo se caracteriza por realizar un uso más completo del Internet, además de lo que ya realizan las personas que ese encuentran en el conglomerado anterior, también realizan actividades más informativas como buscar información sobre salud, sobre bienes y servicios, por otra parte; lo utilizan para enviar y recibir correos, comprar y vender bienes o servicios, realizar trámites bancarios, buscar empleo, gestionar una página web, obtener información de sitios gubernamentales, así como descargar formularios de estos sitios y enviarlos, publicar opiniones y participar en asuntos cívicos o políticos.

En el Cuadro 9 se describen los principales resultados de los conglomerados según las variables sociodemográficas analizadas. Se realizó un análisis de variancia y todas las variables presentan diferencias significativas (ver Anexo 8).

Al analizar la distribución por región de planificación en cada uno de los conglomerados se observa que, en la región Central, la proporción de personas que residen en esta región pertenecen en mayor medida al grupo de ciudadanos digitales. En las restantes regiones se observa una inclinación hacia el grupo de consumidores digitales.

Además, al analizar los perfiles de usuarios de Internet por sexo, se aprecia que los hombres se clasifican mayoritariamente como ciudadanos digitales, esto quiere decir que; hacen un uso más completo del Internet ya que realizan actividades de todo tipo en la red, mientras que las mujeres lo utilizan un poco más para interactuar socialmente, así como para entretenimiento, esto debido a que se encuentran en mayor proporción dentro del grupo de consumidores sociales.

Por grupos de edad se presentan resultados interesantes, las personas entre los 18 y 44 años se categorizan principalmente, como ciudadanos digitales, particularmente para aquellos entre los 25 y 34 años donde se aprecia una notable diferencia, el porcentaje de personas que pertenecen a este grupo es mucho mayor que para aquellos que teniendo la misma edad pertenecen a la categoría de consumidor social. Mientras que las personas de más de 45 años se categorizan como consumidores sociales, principalmente aquellas personas de 55 años o más.

Al analizar los perfiles por nivel educativo se observa que las personas sin ningún nivel de instrucción o con primaria, se ubican dentro del perfil de consumidores sociales, el porcentaje de personas que pertenece al grupo consumidor social es más del doble con respecto a los que se encuentran dentro del perfil ciudadano digital. Para aquellos con nivel de instrucción secundaria esta diferencia se inclina un poco hacia ciudadano digital.

Las personas con nivel de instrucción universitario se encuentran en un porcentaje mucho mayor dentro del grupo ciudadano digital, el triple con respecto al perfil consumidor digital.

En cuanto al nivel de ingresos, se aprecia que aquellas personas que poseen ingresos inferiores a los 250 mil colones se ubican en mayor medida dentro del perfil consumidor social, el porcentaje de consumidores sociales dobla la cifra con respecto a los ciudadanos digitales. Por su parte las personas con un nivel adquisitivo superior a los 500 mil colones corresponden en mayor proporción al perfil ciudadano digital. El porcentaje de personas que pertenece a este perfil y poseen ingresos superiores a los 750 mil es el triple al compararse con el porcentaje de personas del perfil consumidor social.

En resumen, los consumidores sociales en su mayoría son personas de la región Chorotega, Pacífico Central, Huetar Caribe y Huetar Norte; una mayor proporción corresponde a mujeres, a personas de más de 45 años, personas con nivel educativo de primaria o menos, así como personas con ingresos inferiores a los 250 mil colones. Mientras que, los ciudadanos digitales son personas en su mayoría de la región Central, mayoritariamente hombres, con edades entre los 18 y 44 años, personas con nivel de instrucción universitaria y con ingresos superiores a los 750 mil colones.

Cuadro 9. Distribución porcentual de las personas entrevistadas por conglomerado según variables sociodemográficas

		Conglomerados	
		Consumidor social	Ciudadano digital
Región de planificación	Central	21,0	32,2
	Chorotega	15,7	10,6
	Pacífico Central	14,8	13,8
	Brunca	13,5	13,8
	Huetar Caribe	15,8	14,9
	Huetar Norte	19,3	14,7
	Total	100,0	100,0
Sexo	Hombre	43,8	46,9
	Mujer	56,2	53,1
	Total	100,0	100,0

Grupos de edad	18 a 24	19,6	23,5
	25 a 34	22,6	32,5
	35 a 44	17,3	19,7
	45 a 54	19,3	14,7
	55 o más	21,1	9,6
	Total	100,0	100,0
Educación	Primaria o menos	46,1	20,8
	Secundaria	42,8	44,5
	Universitaria	11,0	34,7
	Total	100,0	100,0
Ingreso	Menos de 250 mil colones por mes	42,8	22,4
	Entre 250 mil y menos de 500 mil	39,6	39,8
	Entre 500 mil y menos de 750 mil	10,4	15,4
	Más de 750 mil	7,2	22,4
	Total	100,0	100,0

Nota: Todas las variables presentan diferencias significativas al 5%.

Fuente: MICITT (2017). Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017. [Archivo de datos].

VI. CONCLUSIONES

Al analizar los resultados del índice y sus escalas se manifiestan hallazgos importantes. El índice de uso de Internet presenta una diferencia marcada entre regiones, en la región central se tiene un valor significativamente mayor y en el resto de las regiones, con excepción de la región Brunca la cual no presenta diferencias, es significativamente inferior. En general, la región Central presenta un mayor uso en la escala informativa, de productividad y en la escala de habilitación ciudadana.

El comportamiento del índice por sexo presenta mayor valor para hombres que para mujeres, los hombres manifiestan valores más altos en las escalas de entretenimiento, productividad y habilitación ciudadana, por su parte las mujeres poseen los valores significativamente más bajos justamente en estas escalas.

Existe una clara brecha generacional en cuanto al uso de Internet, las personas entre los 18 y 34 años tienen valores más altos en el índice, conforme aumenta la edad, la media del índice disminuye notablemente. Las personas entre los 18 y 34 años tienen valores más altos en las

escalas de interacción social, productividad y habilitación ciudadana. Las personas entre los 35 y 44 años presentan valores significativamente mayores únicamente en la escala informativa, en el resto de las escalas los valores se encuentran dentro del promedio nacional (escala de interacción social, entretenimiento, productividad y habilitación ciudadana). En los grupos de edad de los 45 años o más, los valores en todas las escalas son significativamente inferiores al promedio nacional.

Se presentan diferencias notables en cuanto al nivel de instrucción, las personas con educación universitaria poseen valores superiores en el índice de uso de Internet, mientras que aquellas con primaria o sin ningún nivel de instrucción poseen valores en el uso significativamente más bajos. Este comportamiento se mantiene en todas las escalas, las personas con estudios universitarios presentan valores significativamente superiores en las cinco escalas, mientras que las personas con primaria o menos poseen valores significativamente más bajos. Existe una relación directa tanto en los valores del índice como en los valores de las escalas, a mayor nivel de instrucción mayor es el valor del indicador.

El nivel de ingreso también es un factor discriminante cuando se habla de uso de Internet, con forme aumenta el poder adquisitivo aumenta también el valor del índice. Las personas con ingresos de más de 500 mil colones tienen valores significativamente más altos y aquellas con ingresos de menos de 250 mil colones presentan un valor en el índice por debajo del promedio nacional. Lo mismo sucede en las cinco escalas analizadas, las personas que poseen ingresos inferiores a los 250 mil colones son las que presentan valores significativamente más bajos y aquellas cuyos ingresos superan los 500 mil colones poseen valores significativamente superiores en todas las escalas.

La creación de perfiles de usuarios permite obtener una caracterización más precisa, se consideran dos grupos, los consumidores sociales y los ciudadanos digitales, encontrándose diferencias entre los perfiles de usuarios. El grupo de ciudadanos digitales pertenecen en mayoría a la región Central, por su parte los consumidores sociales se encuentran en mayor medida en las regiones Chorotega, Pacífico Central, Huetar Caribe y Huetar Norte. Este resultado muestra una notable diferencia y evidencia como en la región Central se concentran

además de personas que realizan un consumo social, usuarios con actividades informativas, de productividad y habilitación ciudadana, mientras que en las restantes regiones se concentran personas que en mayor proporción realizan actividades más sociales y de entretenimiento.

Se encuentran diferencias en los perfiles entre hombres y mujeres, los hombres se categorizan en su mayoría como ciudadanos digitales, realizan un uso de Internet que abarca todas las actividades, informativa, interacción social, entretenimiento, productividad y habilitación ciudadana. Mientras que las mujeres se clasifican como consumidoras más sociales, su actividad en Internet se centra en la interacción social y el entretenimiento.

La edad es un factor importante a la hora de analizar los perfiles de usuarios, las personas más jóvenes, entre los 18 y 44 años se categorizan como ciudadanos digitales, mientras que aquellas de más de 44 años pertenecen en mayor medida al grupo de consumidores sociales. Los resultados muestran la brecha generacional existente en cuanto al uso y apropiación de las TIC, las personas de mayor edad que utilizan Internet lo enfocan en redes sociales, escuchar radio, ver televisión, descargar juegos y aplicaciones, y no tanto en actividades más productivas, informativas o de habilitación y participación ciudadana.

Otra de las variables que presenta diferencias en cuanto a los perfiles de usuarios de Internet es el nivel de educación, para el nivel primaria o menos se tienen personas que pertenecen al perfil consumidor social, el cual se caracteriza por un uso dirigido a redes sociales y entretenimiento. Las personas que poseen estudios universitarios son parte del perfil ciudadano digital, por ende, el uso que hacen de Internet es más integral y abarca todos los componentes del índice.

Con respecto al nivel de ingresos se encuentra que las personas con ingresos inferiores a los 250 mil colones se categorizan principalmente dentro del perfil consumidor social, y aquellos con ingresos superiores a los 500 mil entran mayoritariamente en el perfil ciudadano digital, cabe destacar que las personas con ingresos superiores a los 750 mil colones el porcentaje de personas que pertenecen al perfil ciudadano digital se triplica.

Tal y como lo menciona Sábada (2010), la política pública debería ir orientada a asegurar la inclusión digital especialmente en estos grupos que presentan desigualdad en cuanto al uso de Internet y así evitar que la red se convierta en un medio de separación social, cultural y económica.

Por último, dado que las variables utilizadas son construidas a partir de estándares establecidos por la OCDE y la UIT, la metodología empleada en la investigación permite sistematizar la información y así obtener un comparativo de los países y orientar la política y la toma de decisiones a nivel nacional e internacional.

VII. BIBLIOGRAFÍA

Aldenderfer, M. S. & Blashfield, R. K. (1984). A review of clustering methods. In *Cluster analysis*. SAGE Publications, Inc.

Agudo, S., Pascual, M. Á., & Fombona, J. (2012). Usos de las herramientas digitales entre las personas mayores. *Comunicar*, 20(39), 193-201. <https://doi.org/10.3916/C39-2012-03-10>

Arias, B. (2008). Desarrollo de un ejemplo de análisis factorial confirmatorio con LISREL, AMOS Y SAS. *Metodología en la Investigación sobre discapacidad. Introducción al uso de las ecuaciones estructurales VI Simposio Científico SAID*.

Brown, T. (2006). *Confirmatory Factor Analysis for Applied Research*. New York: The Guilford Press.

Cristóbal-Fransi, E., Daries-Ramon, N., & Baldomar, J. P. (2014). An applied study based in the internet use perspectives. *Cuadernos De Gestión*, 14(1), 33-55. DOI: 10.5295/cdg.120373ec

Fernández, R. (2008). Modelos de medida y análisis factorial confirmatorio. *Metodología en la Investigación sobre discapacidad. Introducción al uso de las ecuaciones estructurales VI Simposio Científico SAID*.

Finch, W. (2020). Introduction to factor analysis. In *Exploratory factor analysis*. SAGE Publications, Inc. <https://www-doi-org.ezproxy.sibdi.ucr.ac.cr/10.4135/9781544339900>

Kaplan, D. (2009). *Statistical Assumptions Underlying Structural Equation Modeling (2.a ed.)*. SAGE Publications, Inc.

Kline, R. (2011). *Principles and Practice of Structural Equation Modeling (3.a ed.)*. The Guilford Press.

Martínez-Abad, F., Rodríguez-Conde, M. (2017). Comportamiento de las correlaciones producto-momento y tetracórica-policórica en escalas ordinales: un estudio de simulación. *Revista Electrónica de Investigación y Evaluación Educativa*, 23(2), 1-21. <http://dx.doi.org/10.7203/relieve.23.2.9476>

Méndez, C., Rondón, M. (2012). Introducción al análisis factorial confirmatorio. *Revista Colombiana de Psiquiatría*, 41(1), 197-207.

Ministerio de Ciencia, Tecnología y Telecomunicaciones (2019). *Acceso y Uso de los Servicios de Telecomunicaciones en Costa Rica 2017*. San José, Costa Rica.

Montero, E. (2008). Escalas o índices para la medición de constructos: el dilemma del analista de datos. *Avances en Medición*, 6(1), 15-24.

Mulaik, S.A. (2009). *Linear causal modeling with structural equations*. New York: CRC Press Taylor & Francis Group.

Nunnally J. & Bernstein I. (1994). *Psychometric theory*. New York: McGraw- Hill.

Orgaz, M. (2008). Introducción a la metodología SEM: Concepto y propósitos fundamentales. *Metodología en la Investigación sobre discapacidad. Introducción al uso de las ecuaciones estructurales VI Simposio Científico SAID*.

Oviedo, H., Campo-Arias, A. (2005). Aproximación al uso del coeficiente alfa de Cronbach. *Revista Colombiana de Psiquiatría*, 24(4), 572-580.

Ruiz, M. (2008). Introducción a los modelos de ecuaciones estructurales. *Metodología en la Investigación sobre discapacidad. Introducción al uso de las ecuaciones estructurales VI Simposio Científico SAID*.

Ruiz, M., Pardo A. & San Martín R. (2010). Modelos de ecuaciones estructurales. *Papeles de Psicólogo*, 31(1), 34-45.

Sedaba, C. (2010). El perfil del usuario de internet en España. *Intervención Psicosocial*, 19(1), 41-55. doi:<http://dx.doi.org.ezproxy.sibdi.ucr.ac.cr:2048/10.5093/in2010v19n1a6>

Superintendencia de Telecomunicaciones (2010-2019). *Estadísticas del Sector de Telecomunicaciones*. <https://sutel.go.cr/informes-indicadores>

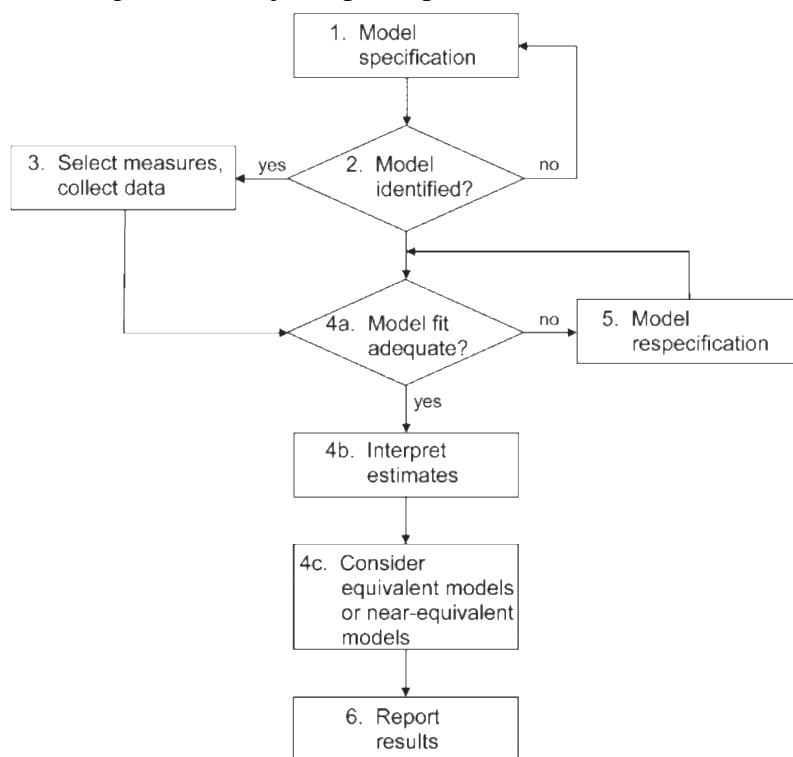
Torres, J., Infante, A. (2011). Desigualdad digital en la universidad: usos de Internet en Ecuador. *Commencer*, 19(37), 81-88. <http://dx.doi.org/10.3916/C37-2011-02-08>

Torres-Díaz, J., Duart, J. M., Gómez-Alvarado, H., Marín-Gutiérrez, I., & Segarra-Faggioni, V. (2016). Internet use and academic success in university Students/Usos de internet y éxito académico en estudiantes universitarios. *Comunicar*, 24(48), 61-69.

Uprichard, E. (2009). Introducing cluster analysis: what can it teach us about the case? In the *SAGE handbook of case-based methods*. SAGE Publications Ltd, <https://www-doi-org.ezproxy.sibdi.ucr.ac.cr/10.4135/9781446249413>

Yang, K. (2010). Statistical case-oriented methods. In *Making sense of statistical methods in social research*. SAGE Publications Ltd, <https://www-doi-org.ezproxy.sibdi.ucr.ac.cr/10.4135/9781473914636>

VIII. ANEXOS

Anexo 1. Diagrama de flujo de pasos para realizar un SEM.

Fuente: Kline. (2011). *Principles and Practice of Structural Equation Modeling*, (p.92) New York: The Guilford Press.

Anexo 2. Media e intervalos de confianza de la escala informativa según variables sociodemográficas

Media nacional		52,8
Intervalos de confianza	Límite superior	55,1
	Límite inferior	50,4
Región de planificación	Central	62,0
	Chorotega	40,0
	Pacífico Central	49,9
	Brunca	54,7
	Huetar Caribe	54,0
	Huetar Norte	46,3
Sexo	Hombre	52,0
	Mujer	53,5
Grupos de edad	18 a 24	53,6
	25 a 34	59,0
	35 a 44	56,0
	45 a 54	49,4

	55 o más	39,9
Educación	Primaria o menos	38,2
	Secundaria	54,0
	Universitaria	71,9
Ingreso	Menos de 250 mil colones por mes	39,9
	Entre 250 mil y menos de 500 mil	53,3
	Entre 500 mil y menos de 750 mil	63,4
	Más de 750 mil	71,5
Ingreso subjetivo	No les alcanza, tienen grandes dificultades	37,5
	No les alcanza, tiene dificultades	43,6
	Les alcanza justo, sin grandes dificultades	54,8
	Les alcanza bien, pueden ahorrar	67,6

Fuente: MICITT (2017). Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017. [Archivo de datos].

Anexo 3. Media e intervalos de confianza de la escala de interacción social según variables sociodemográficas

	Media nacional	39,3
Intervalos de confianza	Límite superior	40,4
	Límite inferior	38,3
Región de planificación	Central	37,5
	Chorotega	44,9
	Pacífico Central	40,9
	Brunca	40,8
	Huetar Caribe	38,0
	Huetar Norte	36,8
Sexo	Hombre	40,2
	Mujer	38,6
Grupos de edad	18 a 24	45,2
	25 a 34	45,2
	35 a 44	39,6
	45 a 54	33,8
	55 o más	26,5
Educación	Primaria o menos	32,1
	Secundaria	40,0
	Universitaria	48,5
Ingreso	Menos de 250 mil colones por mes	33,9
	Entre 250 mil y menos de 500 mil	39,5
	Entre 500 mil y menos de 750 mil	40,6

	Más de 750 mil	48,7
Ingreso subjetivo	No les alcanza, tienen grandes dificultades	38,4
	No les alcanza, tiene dificultades	33,5
	Les alcanza justo, sin grandes dificultades	39,6
	Les alcanza bien, pueden ahorrar	48,3

Fuente: MICITT (2017). Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017. [Archivo de datos].

Anexo 4. Media e intervalos de confianza de la escala de entretenimiento según variables sociodemográficas

	Media nacional	52,4
Intervalos de confianza	Límite superior	53,7
	Límite inferior	51,0
Región de planificación	Central	53,5
	Chorotega	54,9
	Pacífico Central	53,8
	Brunca	51,9
	Huetar Caribe	47,8
	Huetar Norte	51,6
Sexo	Hombre	56,8
	Mujer	48,4
Grupos de edad	18 a 24	63,4
	25 a 34	61,3
	35 a 44	53,7
	45 a 54	42,3
	55 o más	30,8
Educación	Primaria o menos	38,9
	Secundaria	55,3
	Universitaria	66,7
Ingreso	Menos de 250 mil colones por mes	43,4
	Entre 250 mil y menos de 500 mil	53,8
	Entre 500 mil y menos de 750 mil	57,0
	Más de 750 mil	64,7
Ingreso subjetivo	No les alcanza, tienen grandes dificultades	41,5
	No les alcanza, tiene dificultades	45,9
	Les alcanza justo, sin grandes dificultades	54,0
	Les alcanza bien, pueden ahorrar	62,4

Fuente: MICITT (2017). Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017. [Archivo de datos].

Anexo 5. Media e intervalos de confianza de la escala de productividad según variables sociodemográficas

Media nacional		22,0
Intervalos de confianza	Límite superior	23,5
	Límite inferior	20,6
Región de planificación	Central	29,1
	Chorotega	17,1
	Pacífico Central	19,0
	Brunca	18,1
	Huetar Caribe	20,1
	Huetar Norte	21,0
Sexo	Hombre	27,2
	Mujer	17,5
Grupos de edad	18 a 24	27,1
	25 a 34	28,9
	35 a 44	23,5
	45 a 54	15,0
	55 o más	8,8
Educación	Primaria o menos	8,6
	Secundaria	20,9
	Universitaria	43,8
Ingreso	Menos de 250 mil colones por mes	10,1
	Entre 250 mil y menos de 500 mil	20,2
	Entre 500 mil y menos de 750 mil	30,7
	Más de 750 mil	44,0
Ingreso subjetivo	No les alcanza, tienen grandes dificultades	8,8
	No les alcanza, tiene dificultades	13,0
	Les alcanza justo, sin grandes dificultades	23,8
	Les alcanza bien, pueden ahorrar	36,7

Fuente: MICITT (2017). Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017. [Archivo de datos].

Anexo 6. Media e intervalos de confianza de la escala de habilitación ciudadana según variables sociodemográficas

Media nacional		18,5
Intervalos de confianza	Límite superior	19,8
	Límite inferior	17,2
Región de planificación	Central	23,3
	Chorotega	13,6
	Pacífico Central	16,5
	Brunca	19,7

	Huetar Caribe	16,2
	Huetar Norte	16,6
Sexo	Hombre	22,7
	Mujer	14,8
Grupos de edad	18 a 24	20,2
	25 a 34	23,2
	35 a 44	19,8
	45 a 54	15,8
	55 o más	8,9
Educación	Primaria o menos	6,8
	Secundaria	15,4
	Universitaria	41,3
Ingreso	Menos de 250 mil colones por mes	9,1
	Entre 250 mil y menos de 500 mil	16,5
	Entre 500 mil y menos de 750 mil	24,4
	Más de 750 mil	40,2
Ingreso subjetivo	No les alcanza, tienen grandes dificultades	10,8
	No les alcanza, tiene dificultades	10,6
	Les alcanza justo, sin grandes dificultades	20,2
	Les alcanza bien, pueden ahorrar	29,7

Fuente: MICITT (2017). Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017. [Archivo de datos].

Anexo 7. Centros de conglomerados finales para el análisis de dos grupos

	Clúster	
	1	2
Informativa	13,68	88,13
Interacción social	28,59	49,03
Entretenimiento	36,74	66,50
Productividad	7,63	35,09
Habilitación ciudadana	4,61	31,06

Fuente: MICITT (2017). Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017. [Archivo de datos].

Anexo 8. Resultados del análisis de variancia entre las variables sociodemográficas y el análisis de conglomerados de dos grupos

		Suma de cuadrados	gl	Media cuadrática	F	Sig.	Eta
Región * grupo2	Entre grupos (Combinado)	68,889	1,000	68,889	20,476	0,000	0,088
	Dentro de grupos	8788,442	2612,229	3,364			
	Total	8857,332	2613,229				
Género * grupo2	Entre grupos (Combinado)	0,637	1,000	0,637	2,570	0,109	0,031
	Dentro de grupos	647,177	2612,229	0,248			
	Total	647,814	2613,229				
Grupos de edad * grupo2	Entre grupos (Combinado)	134,386	1,000	134,386	73,455	0,000	0,165
	Dentro de grupos	4779,066	2612,229	1,829			
	Total	4913,452	2613,229				
Educación * grupo2	Entre grupos (Combinado)	153,295	1,000	153,295	257,424	0,000	0,300
	Dentro de grupos	1555,574	2612,229	0,595			
	Total	1708,870	2613,229				
Ingreso * grupo2	Entre grupos (Combinado)	121,124	1,000	121,124	37,549	0,000	0,119
	Dentro de grupos	8426,440	2612,229	3,226			
	Total	8547,564	2613,229				

Fuente: MICITT (2017). Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017. [Archivo de datos].

PRÁCTICA PROFESIONAL II

I. INTRODUCCIÓN

Como parte de las acciones que ejecuta el Ministerio de Ciencia, Innovación, Tecnología y Telecomunicaciones (MICITT), para generar información actualizada y periódica, sobre la evolución de las Tecnologías de Información y Comunicación (TIC) en las viviendas y población del país, se realiza la “Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica (EAU)”.

En el país, la apertura del mercado de las telecomunicaciones, en el año 2008, provocó un dinamismo en el sector que beneficia el acceso de más personas a las TIC (SUTEL, 2017), por lo que, en una economía globalizada y en una sociedad cada vez más interconectada, los indicadores y su medición son necesarios.

La EAU brinda insumos que ayudan al sector a velar por el cumplimiento de la normativa vigente, la cual establece en la Ley General de Telecomunicaciones N°8642, la responsabilidad del Estado de garantizar a todos los usuarios el acceso universal y solidario a los servicios de telecomunicaciones, así como la protección de sus derechos en esta materia. Además, esta ley contempla dentro de sus principios rectores el establecimiento de garantías y derechos a favor de los usuarios finales, para que puedan acceder y disfrutar de servicios de calidad, a un precio asequible.

Por lo tanto, la EAU es de suma importancia, ya que brinda información del sector, lo que permite formular y actualizar periódicamente las políticas públicas. También posibilita analizar la información tomando en cuenta diferentes sectores geográficos y establecer acciones diferenciadas para asegurar la inclusión de toda la población a la sociedad de la información y el conocimiento.

La encuesta se realizó anualmente desde el año 2012 hasta el 2018, y para mayo del 2023 se está en proceso de contratación para una actualización de los datos. Desde su primera versión hasta la fecha el contenido del cuestionario ha variado, buscando extraer indicadores

actualizados en materia de telecomunicaciones, utilizando los estándares establecidos por la Organización para la Cooperación y el Desarrollo Económico (OCDE) y por la Unión Internacional de Telecomunicaciones (UIT). Además, debe tomarse en cuenta que, debido al proceso de incorporación del país a la OCDE, la periodicidad de la encuesta es clave para atender la solicitud de indicadores que demanda tanto la OCDE como la UIT. Sumado a esto, debe mencionarse que la situación fiscal que atraviesa el país implica la adopción de medidas de austeridad y contención del gasto público en aras de evitar el desperdicio de recursos. Bajo este contexto se consideró necesario revisar algunas etapas de la EAU, entre las que destaca la determinación del número de Unidades Primarias de Muestreo (UPM) a seleccionar en la muestra para realizar estimaciones con una precisión adecuada y el tamaño del submuestreo o número de entrevistas en cada UPM. Esta revisión es importante porque tiene un impacto directo en los recursos económicos que se destinan a la encuesta.

En la actualidad, no existe en el país un estudio similar que utilice procesos de simulación para evaluar tamaño de muestra de una encuesta, lo cual contribuye metodológicamente para que otras instituciones que realicen estudios cuantitativos puedan hacerlo y determinar si es posible una reducción de la misma en aras de reducir los costos y por ende el gasto público.

El propósito de la evaluación del tamaño de muestra de la EAU es determinar un tamaño óptimo, manteniendo la precisión en las estimaciones. La evaluación se realiza por medio de un proceso de simulación, alternativa metodológica que se está aprovechando en la estadística y que se aplica en diversas ramas de la ciencia.

En el presente documento, inicialmente, se plantean los objetivos de la investigación, seguido de un apartado donde se expone la orientación conceptual en la cual se incluye una revisión de literatura sobre la técnica a utilizar y trabajos similares. En el apartado metodológico se define la fuente de datos, las variables de estudio, la técnica estadística de análisis y el software estadístico utilizado. Luego, se presentan los resultados y el análisis, por último, una sección de conclusiones y de la bibliografía utilizada.

II. OBJETIVOS

2.1 Objetivo general

Calcular el tamaño de muestra o número de entrevistas que deben seleccionarse en la EAU para obtener estimaciones con una precisión determinada en cada una de las regiones de planificación del país y así proveer al MICITT de información pertinente para establecer un tamaño de muestra final.

2.2 Objetivos específicos

1. Determinar el número de UPM que deben seleccionarse en cada una de las regiones de planificación del país para lograr una precisión adecuada.
2. Determinar la cantidad de entrevistas a submuestrear dentro de cada UPM para lograr una precisión adecuada.
3. Desarrollar un proceso de simulación que permita balancear el número de UPM y de entrevistas a submuestrear dentro de cada UPM, para alcanzar una precisión determinada en las estimaciones.

III. ORIENTACIÓN CONCEPTUAL

Los procesos de simulación se han utilizado en diferentes ramas de la ciencia y diferentes estudios muestran la aplicación de la técnica para evaluar el tamaño de muestra en diferentes ámbitos.

Fishman (1971) señala que un objetivo común del análisis estadístico, en un experimento de simulación, es estimar la media de la población partiendo de datos muestrales. En su estudio expone un método para precisar el tamaño de muestra necesario para estimar la media de un proceso, con un nivel específico de precisión estadística y utilizando un experimento de simulación.

Fishman encuentra que conforme aumenta el tamaño de la muestra, el procedimiento arroja resultados que concuerdan más estrechamente con los predichos. Como producto provee un método para determinar el tamaño óptimo de la muestra, estableciendo para el intervalo de confianza un ancho absoluto fijo alrededor de la media.

En este artículo el autor menciona que el procedimiento planteado posee el inconveniente de que las estimaciones obtenidas están directamente relacionadas al tamaño de la muestra elegido previamente.

En el campo de la biología, Bekoff y Mech (1984) realizan simulaciones en el uso del espacio de los animales para determinar la relación entre el rango estimado del área del hogar, su variabilidad y el tamaño de muestra. En esta investigación se entiende el tamaño de muestra como la cantidad de ubicaciones. Los resultados muestran que a medida que aumenta el tamaño de la muestra el rango estimado del área del hogar aumenta asintóticamente, al mismo tiempo que disminuye la variabilidad de las estimaciones. En los resultados se determina el número de ubicaciones para estimar con precisión el área de alcance del hogar.

En la agricultura, Chandra, Huaman, Hari Krishna y Ortiz (2002), usando simulación, desarrollan una estrategia de muestreo óptima con el fin de establecer el tamaño de la colección del núcleo de cultivos. Para representar adecuadamente el espectro genético y capturar al máximo la diversidad genética, los resultados muestran el tamaño de muestra óptimo.

En la industria, Brenner (1965) expone que la simulación se utiliza muchas veces para estimar la manera óptima de operar un sistema de inventario, con el objetivo de minimizar el costo. En su artículo examina cómo el tamaño de muestra de la simulación está relacionado con la sanción económica de elegir algún procedimiento no óptimo para operar el sistema de inventario. El tamaño de muestra se establece considerando el nivel de precisión deseado, la variabilidad del sistema y los recursos disponibles. Se llega a la conclusión de que a mayor tamaño de la muestra menor es el costo económico.

Otra aplicación en este mismo campo la exponen Mendell, Thode y Finch (1991). Estos autores plantean un análisis en la potencia de la prueba y el tamaño de muestra utilizando simulación. El problema de estudio es la mezcla de dos componentes, para lo que definen diferentes rangos de proporciones en la mezcla de ambos componentes. Los resultados muestran que a medida que aumenta el tamaño de la muestra se incrementa también la potencia de la prueba.

Los estudios realizados en el tema proporcionan evidencia para la evaluación del tamaño de muestra utilizando la simulación como una técnica robusta y novedosa en el análisis de datos.

IV. METODOLOGÍA

En este capítulo se describe la fuente de los datos, las variables a utilizar en el análisis, la principal técnica estadística empleada, así como el software estadístico usado para el análisis de los datos.

4.1 Fuente de datos

Los datos utilizados en el estudio corresponden a la “Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017”, realizado por el MICIIT. El trabajo de campo de la encuesta fue realizado por la Unidad de Servicios Estadísticos (USES) de la Escuela de Estadística de la Universidad de Costa Rica (UCR).

La encuesta tiene cobertura nacional y se definen como dominios de estudio las seis regiones de planificación: Central, Chorotega, Pacífico Central, Brunca, Huetar Caribe y Huetar Norte. La población de estudio son todas las personas residentes en Costa Rica, de 18 años o más, que residen habitualmente en viviendas particulares a noviembre de 2017. La unidad informante definida en la encuesta corresponde a toda persona residente en el país de 18 años o más. Para la selección del informante se utilizó el criterio de la persona más próxima a cumplir años (MICITT, 2019).

El diseño muestral es estratificado, bietápico, mediante selección con probabilidad proporcional al tamaño (PPT), utilizando como medida de tamaño el total de viviendas de

cada UPM. La muestra planeada fue de 3 500 entrevistas y se asignaron 1 000 a la Región Central y 500 en cada una de las cinco regiones restantes. Para la selección de la muestra se utilizó el Marco Muestral de Viviendas 2011 (MMV-2011) del Instituto Nacional de Estadística y Censos (INEC).

4.2 Variables de estudio

En el Cuadro 10 se presenta la distribución porcentual de las variables sociodemográficas. Cabe destacar que a la base de datos se le aplica una corrección por no respuesta y se toma como base la ENAHO 2017 para realizar el ajuste por sexo y grupos de edad.

Cuadro 10. Distribución porcentual de las variables sociodemográficas la EAU 2017

Variables	Frecuencia	%
Total	3 618	100
Región		
Central	1 006	27,8
Chorotega	492	13,6
Pacífico Central	520	14,4
Brunca	495	13,7
Huetar Atlántico	508	14,0
Huetar Norte	598	16,5
Sexo		
Masculino	1 729	47,8
Femenino	1 889	52,2
Grupos de edad		
18 a 24	586	16,2
25 a 34	755	20,9
35 a 44	641	17,7
45 a 54	610	16,9
55 o más	1 026	28,3
Educación		
Primaria y menos	1 644	45,4
Secundaria	1 331	36,8
Universitaria	609	16,8

NS/NR	6	0,2
Ingresos		
Menos de 250 mil colones	1 438	39,8
Entre 250 mil y menos de 500 mil	1 229	34,0
Entre 500 mil y menos de 750 mil	363	10,0
Más de 750 mil	409	11,3
NS/NR	179	4,9
Ingresos subjetivos		
No les alcanza, tienen grandes dificultades	340	9,4
No les alcanza, tiene dificultades	1 197	33,1
Les alcanza justo, sin grandes dificultades	1 520	42,0
Les alcanza bien, pueden ahorrar	530	14,7
NS/NR	31	0,9

Fuente: MICITT. Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017.

La encuesta, entre otros, posee el módulo de usos de Internet que cuenta con una batería de 33 preguntas (Anexo 1) y se consulta si ha utilizado o no Internet para realizar diferentes actividades. Con esta información se construyeron las escalas de uso de Internet y un índice de uso de Internet, que corresponden a las principales variables de estudio. Estas variables fueron construidas como parte del trabajo realizado en la Práctica Profesional I, denominada “Caracterización de los usuarios a partir de escalas de uso de Internet” (Aguilar, 2021). Estas escalas y el índice son básicas, porque forman parte de una propuesta ante los organismos internacionales para aplicar la misma metodología en los países miembros y obtener comparabilidad entre ellos. Por este motivo, la precisión en la estimación de estas variables es sumamente importante. Las variables que fueron utilizadas son:

- Índice de uso de Internet (mide el uso global de diferentes actividades que realizan las personas en Internet).
- Escala informativa (mide en qué grado las personas utilizan Internet para obtener información de bienes, servicios y salud).
- Escala de interacción social (mide en qué grado se utiliza Internet para crear contenido e interactuar entre amigos(as) o conocidos).
- Escala de entretenimiento (mide en qué grado las personas utilizan Internet para ocio o entretenimiento).

- Escala de productividad (mide en qué grado las personas utilizan Internet en el ámbito laboral, educativo, personal, profesional, venta o compra de bienes y servicios).
- Escala de habilitación ciudadana (mide en qué grado las personas utilizan Internet para obtener información gubernamental y participación de asuntos cívicos y políticos).

Tanto el índice como las escalas oscilan entre 0 y 10, siendo 0 el equivalente a ningún uso de Internet y 10 es el uso máximo. En el Cuadro 11 se muestra para el índice y las escalas de uso de Internet la media y el intervalo de confianza, tanto a nivel nacional como para cada una de las regiones de planificación. El cálculo del intervalo se hizo tomando en consideración el diseño muestral complejo de la EAU.

A nivel nacional, el índice de uso de Internet tiene una media de 3,2 y entre las regiones la Central es la que presenta el mayor valor (media de 3,6). La escala informativa tiene una media nacional de 5,3 y también es la región Central la que presenta una media más alta (6,2). Por otro lado, la escala de interacción social tiene una media en el ámbito nacional de 3,9 y por regiones la media más alta se encuentra en las regiones Chorotega, Pacífico Central y Brunca (4,5, 4,1 y 4,1 respectivamente). La escala de entretenimiento tiene una media nacional de 5,2 y el comportamiento entre las regiones es muy similar, pues oscila entre 4,8 y 5,5. Con respecto a la escala de productividad y habilitación ciudadana, la media nacional es de 2,2 y 1,8, respectivamente, y para ambas la región Central presenta una media superior a las demás, incluso superando a la media nacional (2,9 y 2,3, respectivamente).

Cuadro 11. Media e intervalos de confianza del índice y las escalas de uso de Internet en el país según región de planificación

Indicador	Región	Media	Intervalo de confianza	
			Límite inferior	Límite superior
Índice de uso de Internet	Costa Rica	3,2	3,1	3,3
	Central	3,6	3,4	3,8
	Chorotega	3,0	2,7	3,2
	Pacífico Central	3,1	2,7	3,4
	Brunca	3,1	2,8	3,3

	Huetar Caribe	3,0	2,7	3,3
	Huetar Norte	3,0	2,6	3,4
	Costa Rica	5,3	5,0	5,5
	Central	6,2	5,8	6,6
	Chorotega	4,0	3,4	4,6
Escala informativa	Pacífico	5,0	4,2	5,8
	Central	5,5	5,0	5,9
	Brunca	5,4	4,8	6,0
	Huetar Caribe	4,6	3,9	5,4
	Huetar Norte	4,6	3,9	5,4
	Costa Rica	3,9	3,8	4,0
	Central	3,7	3,5	3,9
	Chorotega	4,5	4,3	4,7
Escala de interacción social	Pacífico	4,1	3,8	4,4
	Central	4,1	3,8	4,4
	Brunca	4,1	3,8	4,4
	Huetar Caribe	3,8	3,6	4,0
	Huetar Norte	3,7	3,4	4,0
	Costa Rica	5,2	5,1	5,4
	Central	5,4	5,1	5,6
	Chorotega	5,5	5,1	5,8
Escala de entretenimiento	Pacífico	5,4	4,9	5,9
	Central	5,2	4,8	5,6
	Brunca	4,8	4,5	5,0
	Huetar Caribe	5,2	4,8	5,5
	Huetar Norte	5,2	4,8	5,5
	Costa Rica	2,2	2,1	2,3
	Central	2,9	2,7	3,2
	Chorotega	1,7	1,4	2,0
Escala de productividad	Pacífico	1,9	1,5	2,3
	Central	1,8	1,5	2,1
	Brunca	2,0	1,6	2,4
	Huetar Caribe	2,1	1,6	2,6
	Huetar Norte	2,1	1,6	2,6
	Costa Rica	1,8	1,7	2,0
	Central	2,3	2,1	2,6
	Chorotega	1,4	1,1	1,6
Escala de habilitación ciudadana	Pacífico	1,7	1,3	2,0
	Central	2,0	1,7	2,2
	Brunca	1,6	1,3	2,0

Huetar Norte	1,7	1,3	2,0
--------------	-----	-----	-----

Fuente: MICITT (2017). Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017. [Archivo de datos].

4.3 Técnicas estadísticas

Se desarrolló un proceso de simulación en dos etapas para evaluar el tamaño de muestra de la EAU: en la primera etapa se determinó el número de UPM a seleccionar y en la segunda el tamaño del submuestreo o número de entrevistas a realizar en cada UPM.

Desde un punto de vista estadístico, los procesos de simulación serán usados en este estudio para modelar y analizar sistemas complejos, usando modelos matemáticos (Bú, 1993). En este proceso pueden ajustarse parámetros (modificando el número de UPM o tamaño del submuestreo) lo que origina diferentes escenarios para analizar los cambios experimentados. Tal y como lo afirman Robert y Casellas (2010), sobre el uso de la simulación en el muestreo, esta será aprovechada para estimar cantidades desconocidas (número de UPM o tamaño del submuestreo) o realizar nuevas inferencias (intervalos de confianza para determinar precisión).

Más concretamente, bajo esta perspectiva, la evaluación del tamaño de la muestra consiste, como lo afirma Kline (2011), en simular datos con un modelo, utilizando diferentes tamaños de muestra, ajustando el modelo y evaluando la precisión y estabilidad de los resultados. La evaluación consiste en analizar la precisión y estabilidad de los resultados después de aplicar los modelos de simulación (Lenth, 2011). Esto se logra calculando medidas de precisión, tales como los errores estándar o los intervalos de confianza. De acuerdo con Krejcie y Morgan (1970), para lograr lo anterior, se debe establecer el nivel de confianza con el que se desean los resultados y especificar un margen de error máximo para las estimaciones. Por lo general se utiliza un nivel de confianza del 95%. Estos mismos autores señalan que el diseño de la investigación y los objetivos específicos deben tomarse en cuenta para determinar el tamaño de muestra óptimo.

4.3.1 Evaluación de la cantidad de UPM

Con el fin de evaluar el tamaño de muestra de la EAU, se establecen diferentes escenarios, a partir de variaciones en la cantidad de UPM, en cada región de planificación, revisando la precisión de las estimaciones y tomando en consideración el diseño muestral complejo de la EAU. Se toma en cuenta la teoría del muestreo para considerar el diseño muestral complejo de la encuesta, tanto en lo referente a la selección de la muestra como al cálculo de la precisión de las estimaciones.

La idea del análisis era establecer un criterio para seleccionar el número de UPM y, posteriormente, de submuestra en cada UPM, con el fin de mejorar la planificación de la encuesta, tomando en cuenta factores que pueden estar asociados tanto a la selección de la muestra como al trabajo de campo. En una primera etapa se determinó la cantidad de UPM y en la segunda el tamaño del submuestreo.

El proceso para determinar la cantidad de UPM se realizó en fases, de la siguiente manera:

1. Para una región específica, se contabilizaron las UPM disponibles en el archivo de datos de la encuesta (Cuadro 12).
2. Se extrajeron muestras aleatorias de UPM, sin exceder el máximo disponible de UPM de esa región.
3. Para desarrollar los escenarios, se inició con 5 UPM y se tomó en cuenta el máximo de UPM existente en cada región. A final de cuentas, para cada región se desarrollaron 20 escenarios, generados de acuerdo con el número de UPM de cada región.
4. El procedimiento anterior se realizó 5 000 veces ($K=5\ 000$ iteraciones).
5. Para cada escenario y en cada iteración se calculó la media para cada una de las variables de estudio. Es decir, para una cantidad de 5 UPM se calcularon 5 000 medias de cada una de las variables, luego, para la siguiente cantidad de UPM, por ejemplo 8, se volvieron a calcular 5 000 medias de la variable y así sucesivamente.

6. Para cada uno de los 20 escenarios se calcularon los cuantiles 2,5 y 97,5, lo que determinó el intervalo de confianza, y la diferencia entre sus límites determinó el rango.
7. Como criterio para establecer la cantidad óptima de UPM se definió que el rango no debía sobrepasar el valor de 0,5 puntos. Se utilizó este criterio debido a que se deseaba una precisión suficientemente alta y utilizando este valor el error permitido en la estimación era bajo y se ajustaba con la teoría y las recomendaciones de los autores mencionados.
8. El procedimiento de simulación descrito se repitió en cada una de las seis regiones de planificación y para cada una de las seis variables de estudio.

La generación de las muestras de UPM se realizó de manera aleatoria y los resultados se presentan mediante visualizaciones gráficas.

4.3.2 Evaluación del tamaño de la submuestra

Una vez que se determinó la cantidad de UPM en cada región de planificación, se establecieron diferentes escenarios variando la submuestra, es decir, la cantidad de entrevistas en cada UPM. Esto se hizo de dos maneras: la primera consistía en tomar como base el número de UPM óptimo que se obtuvo de la etapa anterior para determinar, en esta etapa, el tamaño de submuestra. La segunda fue utilizar el número de UPM realizado en la EUA y calcular en esta etapa la submuestra por UPM.

Tomando en consideración lo anterior, el procedimiento seguido fue el siguiente:

1. Para una región específica, para cada UPM, se escogieron diferentes tamaños de submuestra, iniciando con una de 3 entrevistas, hasta alcanzar el máximo disponible en cada UPM. En promedio cada UPM tenía 18 entrevistas, por lo que el tamaño máximo de submuestra establecido fue de 17. En cada región se generaron 9 escenarios.
2. El procedimiento anterior se realizó 5 000 veces ($K=5\ 000$ iteraciones).

3. Para cada escenario y en cada iteración se calculó la media y desviación de la variable de estudio. Es decir, para una cantidad de 3 entrevistas se sacaron 5 000 UPM y se calcularon 5 000 medias de la variable de interés, luego, para la siguiente cantidad de entrevistas y así sucesivamente.
4. Para cada uno de los escenarios o tamaños de submuestra, se calcularon los cuantiles 2,5 y 97,5 (para tener los intervalos de confianza) y la diferencia entre los límites fue el rango.
5. Como criterio para establecer el submuestreo óptimo se definió que el rango no sobrepasara el valor de 0,5 puntos, mismo que se utilizó en la etapa anterior.
6. El procedimiento se realizó en cada una de las regiones de planificación y para las variables de estudio que se consideraron pertinentes.

Igualmente, se revisó la precisión en las estimaciones considerando el diseño muestral de la EAU. Los resultados se presentan por medio de visualizaciones gráficas.

Cuadro 12. Distribución del tamaño de muestra y cantidad de UPM según región de planificación

Región de planificación	Tamaño de muestra	Cantidad de UPM de la muestra	Promedio de entrevistas por UPM
Central	1 006	60	17
Chorotega	492	28	18
Pacífico Central	520	27	19
Brunca	495	29	17
Huetar Caribe	508	27	19
Huetar Norte	598	32	19
Total	3 618	203	18

Fuente: MICITT (2017). Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017. [Archivo de datos].

4.3.3 Resumen de escenarios desarrollados en el proceso de simulación

En el Cuadro 13 se presenta un resumen de los procesos de simulación ejecutados en las dos etapas.

Cuadro 13. Procesos de simulación utilizados para evaluar el tamaño de muestra de la EAU en cada una de las variables

Región	UPM		SUBMUESTRA				Totales
	Muestra original	Simulaciones para el número de UPM	Simulaciones para submuestra manteniendo número original de UPM en la muestra		Simulaciones para submuestra con número óptimo de UPM		
	Cantidad UPM	Escenarios	Cantidad UPM	Escenarios	Cantidad UPM	Escenarios	
Central	60	20	60	9	24	9	
Chorotega	28	20	28	9	18	9	
Pacífico Central	27	20	27	9	27	9	
Brunca	29	20	29	9	19	9	
Huetar Caribe	27	20	27	9	21	9	
Huetar Norte	32	20	32	9	32	9	
Total de escenarios		120		54		54	228
Total de iteraciones (k= 5 000)		600 000		270 000		270 000	1 140 000

En la primera fase, para determinar el número de UPM, se simularon 20 escenarios o cantidades de UPM en cada una de las regiones. En cada región, cada uno de los 20 escenarios quedó determinado por la cantidad de UPM generadas aleatoriamente. Así, por ejemplo, en la región Central, donde había 60 UPM, se dividió $60-5 \text{ UPM}/20 \text{ escenarios}$. El valor 5 surge porque es el valor mínimo de UPM que puede existir en cada escenario y el resultado de esa operación fue el espaciamiento aplicado en esta región para establecer la cantidad de UPM que tendría cada escenario en la región. En este caso de la región Central, la cantidad de UPM para cada uno de los 20 escenarios fueron, específicamente, de 5, 8, 11, 13, 16, 19, 22, 24, 27, 30, 33, 35, 38, 41, 44, 46, 49, 52, 55 y 57 UPM. En las otras regiones se procedió de manera similar.

En total se simulan, para esta fase 120 escenarios y como cada uno se iteró 5 000 veces esto dio como resultado un procesamiento de 600 000 datos aleatorios. En un computador estándar este proceso tarda aproximadamente 6 horas. Cabe destacar que este proceso se realiza para cada una de las 6 variables de estudio, por lo tanto, los datos que se presentan en la cantidad de escenarios, iteraciones y tiempo computador hay que multiplicarlo por 6.

En la segunda etapa, para determinar el tamaño del submuestreo o número de entrevistas a realizar en cada UPM, se realizaron 9 escenarios, tanto en las simulaciones donde se mantuvo la cantidad de UPM de la muestra original, como cuando se utilizó la cantidad óptima de UPM. En este caso, para todas las regiones, se utilizaron los siguientes escenarios con un tamaño de submuestra de 3, 4, 6, 8, 10, 12, 14, 16 y 17 entrevistas. En total se simulan 54 escenarios para la primera alternativa, cada uno se iteró 5 000 veces y esto dio como resultado un procesamiento de 270 000 datos. Esto significa que el procesamiento para obtener la cantidad de submuestra suma un total de 540 000 datos simulados para las dos alternativas consideradas. En un computador estándar se tarda aproximadamente 6 horas. A diferencia de la etapa anterior, esta etapa se realiza una única vez para la variable del índice de uso de Internet.

4.4 Software estadístico

Para el análisis descriptivo se utilizó el programa estadístico de análisis de datos, *SPSS (Statistical Package for the Social Sciences)* versión 22. En el proceso de simulación se utilizó el lenguaje de programación R (versión 4.2.1).

V. RESULTADOS

La evaluación del tamaño de muestra se realizó en dos fases: la primera establece la cantidad de UPM que requiere la encuesta y la segunda define el tamaño de la submuestra, según se explicó en las secciones 4.3.1 y 4.3.2. El procedimiento se realizó para cada región de planificación con las estimaciones del índice de uso de Internet y, posteriormente, con las escalas.

La valoración en función de la variable del índice de uso de Internet es primordial, ya que es la variable que engloba a las escalas de uso de Internet, y aunque estas también tienen un uso importante en los resultados, el índice sin duda será el principal criterio en caso de requerir algún recorte a nivel presupuestario.

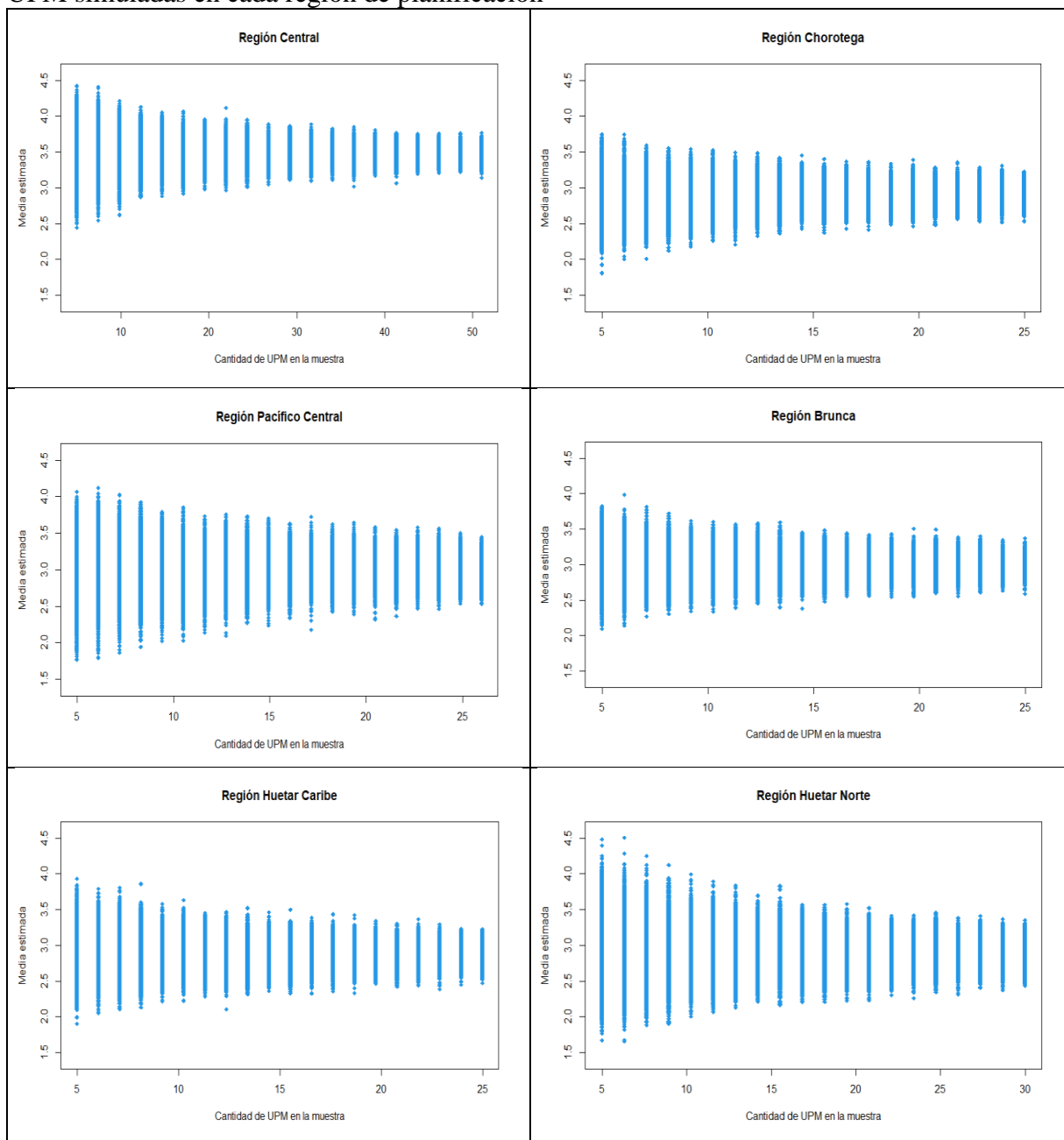
5.1 Número de UPM en la muestra

En esta primera fase se evalúa la cantidad de UPM en función de las estimaciones del índice de uso de Internet.

La Figura 5 resume la simulación del comportamiento de las medias del índice de uso de Internet para las diferentes cantidades de UPM. Es decir, esta figura contiene, para cada cantidad de UPM seleccionadas, las 5 000 medias del índice. Esto se hace en cada una de las regiones de planificación.

Como es de esperar, a medida que aumenta el número simulado de UPM, la variabilidad de las medias simuladas tiende a decrecer. Las particularidades de cada región revelan varios detalles. En la Central, región que contiene 60 UPM, la variabilidad de las medias no decrece mucho más a partir de las 30 UPM, lo que evidencia una posible reducción de la muestra. En la Chorotega, donde hay 28 UPM, la variabilidad de las medias se estabiliza a partir de 20, por lo que también existen posibilidades de reducción. En la región Pacífico Central, la que posee un total de 27 UPM, aún con 25 UPM la variabilidad de las medias no disminuye considerablemente, lo que tentativamente sugiere un aumento en la cantidad de UPM. En las regiones Brunca y Huetar Caribe, donde hay 29 y 27 UPM, respectivamente, a partir de 20 la variabilidad de las medias dejar de disminuir. La región Huetar Norte, la cual tiene un total de 32 UPM, no solo presenta una alta variabilidad, sino que esta disminuye poco conforme aumenta la cantidad de UPM, lo que hace suponer que requiere más cantidad de UPM.

Figura 5. Comportamiento de las medias del índice de uso de Internet según número de UPM simuladas en cada región de planificación



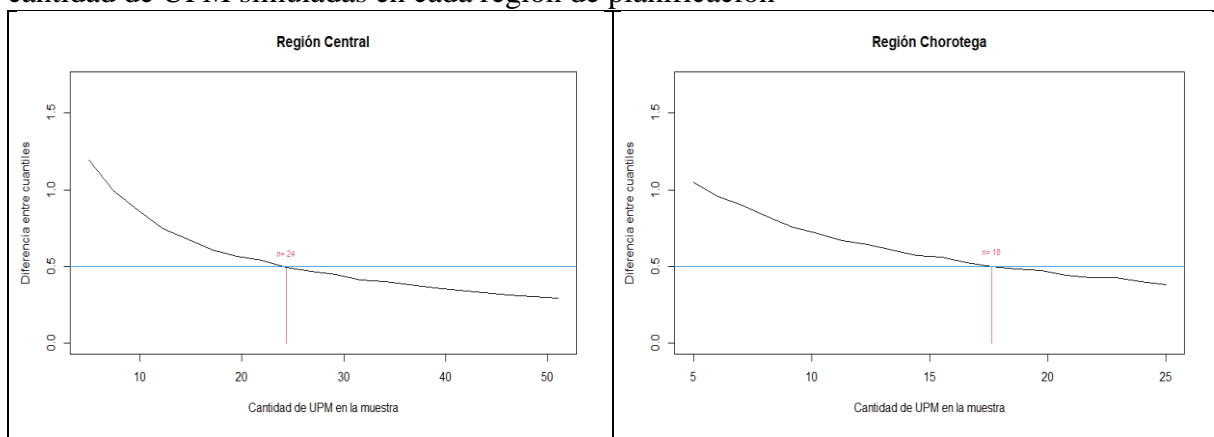
Fuente: MICITT. Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017.

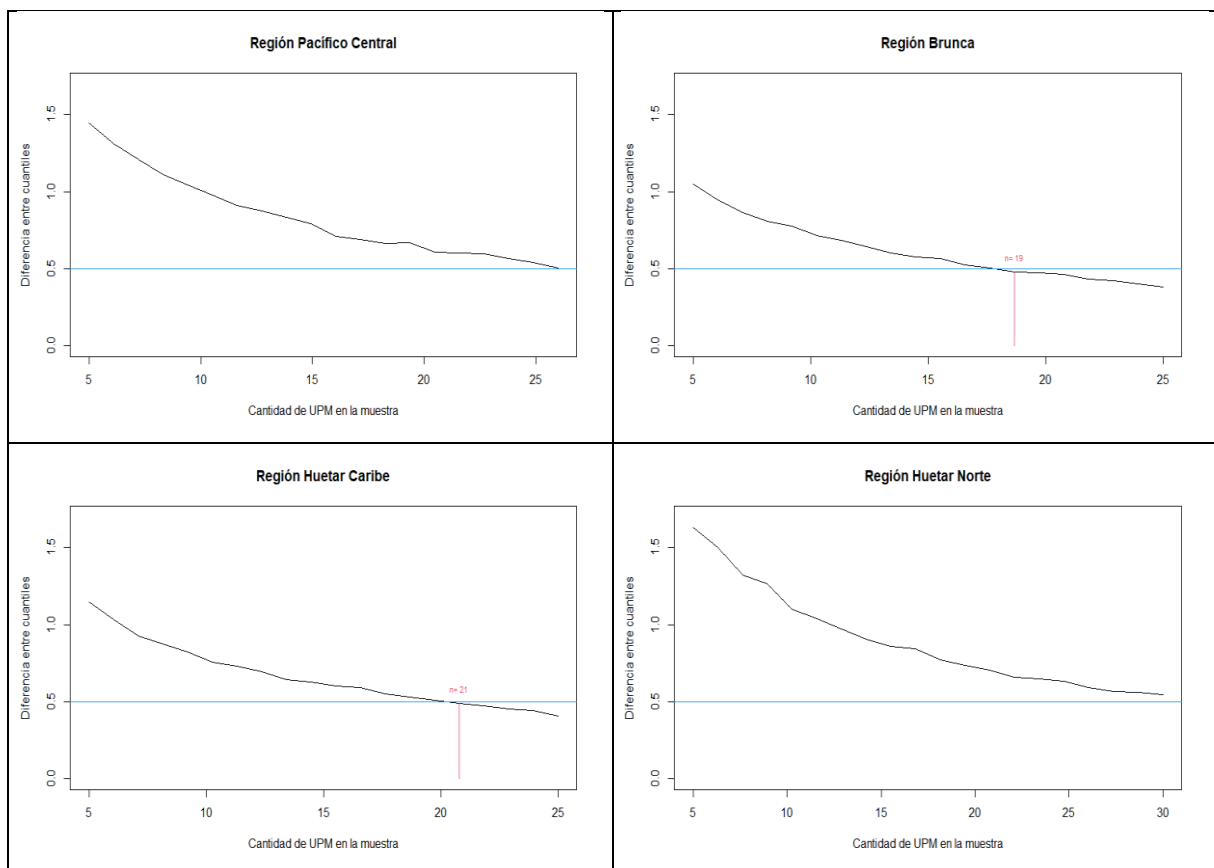
Como complemento, en la Figura 6 se observa, para cada número simulado de UPM, la diferencia entre los cuantiles de las medias simuladas del índice de uso de Internet en cada una de las regiones. El criterio utilizado para establecer el número de UPM deseado en la muestra es que la diferencia entre los cuantiles 2,5 y 9,75 de las medias simuladas no sobrepase el valor de 0,5 puntos.

En la región Central, a partir 24 UPM, la diferencia entre los cuantiles presenta una amplitud menor a la establecida, por lo que la cantidad puede reducirse a menos de la mitad. En la Chorotega el criterio usado se cumple con 18 UPM, lo que implica una reducción de 10 UPM. En las regiones Pacífico Central y Huetar Norte no se alcanza el criterio, pese a ello, es factible verificar que se encuentra muy cerca de alcanzarlo, por lo que puede mantenerse la cantidad de UPM actuales. En la región Brunca el criterio se cumple para una cantidad de 19 UPM, resultado que evidencia una reducción considerable en el número actual de UPM en la muestra. En la región Huetar Caribe la diferencia entre los cuantiles llega al límite propuesto para una cantidad de 21 UPM, por lo que se reduce en 6 la cantidad de UPM con respecto al número actual.

Después de realizar este análisis no se encuentra evidencia para aumentar la cantidad de UPM actual en la muestra, ya que en los casos cuando el criterio establecido no se cumple, es por muy poco, por lo que es factible mantener el número de UPM en estas regiones sin que se pierda precisión en el cálculo del índice de uso de Internet de una manera considerable en cada región.

Figura 6. Diferencia entre los cuantiles de las medias del índice de uso de Internet según cantidad de UPM simuladas en cada región de planificación





Fuente: MICITT. Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017.

El Cuadro 14 resume los resultados del tamaño de muestra tomando en consideración las estimaciones del índice de uso de Internet. En la región Central el número de UPM puede reducirse de 60 a 24 UPM (36 UPM menos), lo que representa una disminución en el tamaño de muestra de 603 entrevistas, si se mantiene el submuestreo actual de 17 entrevistas por UPM. En las regiones Chorotega, Brunca y Huetar Caribe el número de UPM se reduce respectivamente en 10, 10 y 6. La muestra total final pasa de 3 618 a 2 556, lo que se traduce en una reducción total de 1 062 entrevistas.

Cuadro 14. Distribución del tamaño de muestra y cantidad de UPM al evaluar las estimaciones del índice de uso de Internet según región de planificación

Región de planificación	Cantidad de UPM de la muestra	Tamaño de muestra	Cantidad de UPM obtenidas en la simulación	Tamaño de muestra propuesto
Central	60	1 006	24	402
Chorotega	28	492	18	316
Pacífico Central ^{1/}	27	520	27	520
Brunca	29	495	19	324
Huetar Caribe	27	508	21	395
Huetar Norte ^{1/}	32	598	32	598
Total	203	3 618	141	2 556

^{1/}Se mantiene la cantidad de UPM asumiendo un error un poco más grande.

Fuente: MICITT. Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017.

Al evaluar el tamaño de muestra considerando las estimaciones de las escalas de uso de Internet, se tienen diferentes escenarios (Cuadro 15). En las estimaciones de la escala informativa no se alcanza el criterio para la reducción de la cantidad de UPM en ninguna de las regiones (se ilustra gráficamente en Anexo 2 y Anexo 3). Cabe destacar que este indicador está compuesto únicamente por dos variables, por lo que su comportamiento podría ser inestable. Además, de las escalas analizadas, esta es la que presenta mayor variabilidad, pese a que su construcción se sustenta desde la teoría, por lo que el análisis se realiza con las otras escalas y en la elección de la cantidad de UPM no se toma en cuenta y se mantiene la cantidad actual.

En la escala de interacción social se cumple el criterio para todas las regiones (Anexo 4 y 5), arrojando una reducción importante en la cantidad de UPM. Por otra parte, la escala de entretenimiento (Anexo 6 y 7) alcanza el criterio en la región Central y Huetar Caribe. Para las demás regiones el criterio no se cumple, aunque en el Anexo 7 se aprecia gráficamente que se encuentra muy cerca el límite establecido. En el caso de la escala de productividad (Anexo 8 y 9) y de habilitación ciudadana (Anexo 10 y 11), el criterio se cumple en la región Central, Chorotega y Brunca, y en los anexos 9 y 11 se aprecia que la diferencia entre los cuantiles se encuentra muy cerca del límite establecido.

Cuadro 15. Distribución la cantidad de UPM al evaluar las estimaciones de las escalas de uso de Internet según región de planificación

Región de planificación	Cantidad de UPM de la muestra	Cantidad de UPM obtenidas en la simulación					
		Índice de uso de Internet	Escala informativa	Escala de interacción social	Escala de entretenimiento	Escala de productividad	Escala de habilitación ciudadana
Central	60	24	n.a.	24	32	39	29
Chorotega	28	18	n.a.	14	n.a.	24	21
Pacífico Central	27	n.a.	n.a.	19	n.a.	n.a.	n.a.
Brunca	29	19	n.a.	22	n.a.	23	17
Huetar Caribe	27	21	n.a.	16	18	n.a.	n.a.
Huetar Norte	32	n.a.	n.a.	29	n.a.	n.a.	n.a.

Nota: Se indica No aplica (n.a.) cuando no alcanza el criterio para determinar un número menor de UPM
Fuente: MICITT. Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017.

Al incluir las estimaciones de las escalas de uso de Internet y analizar los resultados por región de planificación, se obtiene una cantidad de UPM de 39 para la región Central, ya que según el análisis es la cantidad de UPM de mayor tamaño requerida entre las variables analizadas. En la región Chorotega y Brunca la cantidad de UPM queda en 28 y 29 respectivamente, tal y como se encuentra en la muestra original. Cabe mencionar que para ambas regiones la mayoría de las escalas cumplió el criterio, lo que evidencia una posible reducción en la cantidad de UPM, exceptuando a la escala de entretenimiento. Pese a esto, en el Anexo 7 se observa que para estas regiones el criterio estuvo muy próximo a cumplirse, pues casi roza la línea límite que determina la precisión deseada, por lo que se sugiere mantener la cantidad de UPM.

En la región Huetar Caribe también se mantiene la cantidad original, de 27 UPM, ya que las escalas de productividad y habilitación ciudadana, aunque no cumplen el criterio, en los anexos 9 y 11 se aprecia gráficamente que la diferencia entre los cuantiles prácticamente se cumple con la cantidad de UPM de la muestra original, ya que la línea está muy próxima al límite establecido.

Finalmente, en las regiones Pacífico Central y Huetar Norte únicamente se alcanza el criterio para la escala de interacción social. Aunque en los anexos 7, 9 y 11 se aprecia que el criterio

se encuentra próximo a cumplir en las escalas de las diferentes regiones, puede valorarse un pequeño ajuste en la cantidad de UPM en cada región. Por este motivo se considera que con 5 UPM adicionales en cada región el criterio se ajustaría para las demás escalas de uso de Internet.

En síntesis, en el Cuadro 16 y con base en el análisis anterior, se describen los resultados del número óptimo de UPM en la muestra. Obsérvese que al evaluar la cantidad de UPM, utilizando únicamente la variable del índice de uso de Internet hay una disminución de 62 UPM, y al incluir las variables de las escalas de uso de Internet disminuye en 11.

Cuadro 16. Cantidad de UPM de la muestra original y cantidad de UPM óptimo

Región de planificación	Cantidad de UPM original	Cantidad de UPM óptimo únicamente utilizando la variable del índice de uso de Internet	Cantidad de UPM óptimo incluyendo las variables de las escalas de uso de Internet
Central	60	24	39
Chorotega	28	18	28
Pacífico Central	27	27	32
Brunca	29	19	29
Huetar Caribe	27	21	27
Huetar Norte	32	32	37
Total	203	141	192

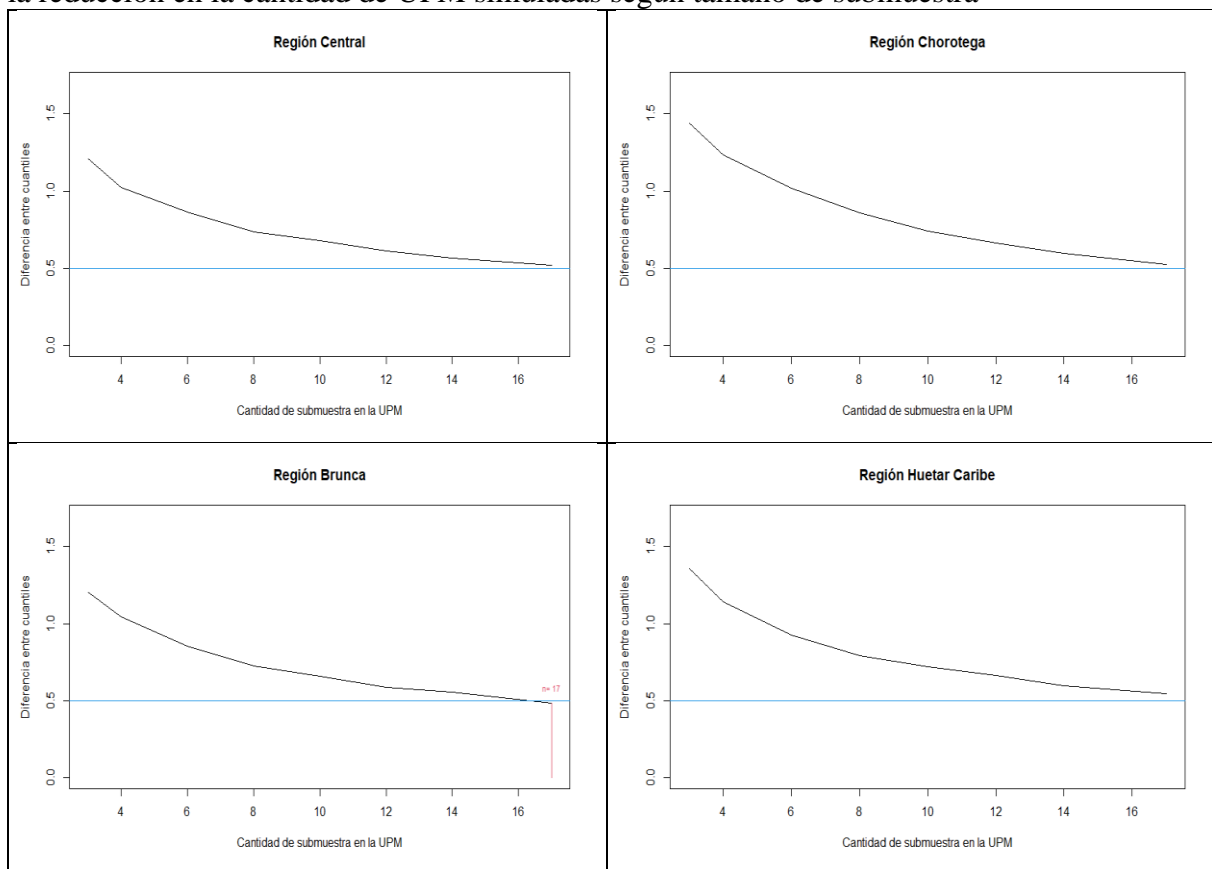
5.2 Cantidad de entrevistas en cada UPM

El análisis de la submuestra por UPM se realizó evaluando únicamente las estimaciones del índice de uso de Internet.

El primer análisis considera la cantidad óptima de UPM obtenidas en el proceso de simulación anterior. En la Figura 7 se tiene, para cada tamaño de submuestra, la diferencia entre los cuantiles de las medias simuladas del índice de uso de Internet en cada una de las regiones donde hubo una reducción en la cantidad de UPM (en la región Pacífico Central y Huetar Norte no hubo disminución en la cantidad de UPM). Obsérvese que el criterio está muy cerca de cumplirse en la cantidad máxima de submuestra, lo que sugiere mantener la

submuestra original (en promedio 18 entrevistas). Esto evidencia que al realizar una reducción en la cantidad de UPM para las regiones correspondientes hace que no sea viable una reducción de la submuestra.

Figura 7. Diferencia entre los cuantiles de las medias del índice de uso de Internet utilizando la reducción en la cantidad de UPM simuladas según tamaño de submuestra



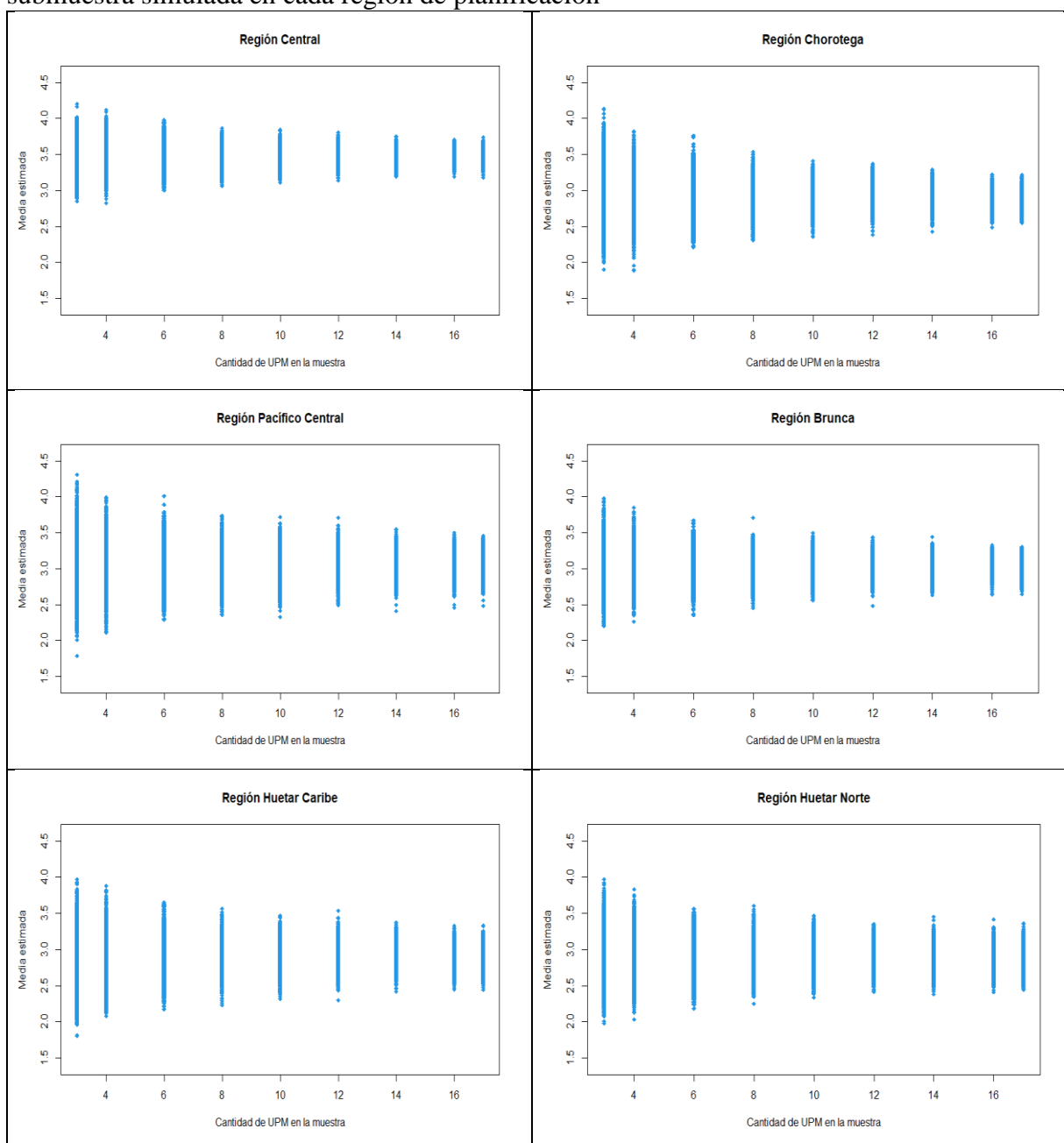
Nota: En la región Central se utiliza 24 UPM, en la región Chorotega 18 UPM, en la región Brunca 19 UPM y en la región Huetar Caribe 21 UPM.

Fuente: MICITT. Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017.

Para analizar otra alternativa, el análisis de la submuestra por UPM se realizó tomando en cuenta la cantidad original de UPM de la muestra. En la Figura 8 se tiene el comportamiento de las medias del índice para las cantidades de submuestra analizadas. Obsérvese que, en la región Central, con un total de 60 UPM, la variabilidad de las medias se estabiliza a partir de la submuestra de tamaño 12, lo que hace suponer una disminución en la cantidad de entrevistas. En la región Chorotega y Pacífico Central, con el número original de UPM, la variabilidad de las medias tiende a mantenerse en los niveles esperados en la submuestra de tamaño 16, por lo tanto, se podría esperar una leve disminución en la cantidad de entrevistas.

En la región Brunca se aprecia que la variabilidad de las medias no decrece mucho más a partir de 14 entrevistas, lo que evidencia una posible reducción en la submuestra por UPM. Para las regiones Huetar Caribe y Huetar Norte las medias se estabilizan a partir de 16 entrevistas, por lo tanto, se esperaría que la submuestra disminuya levemente o se mantenga en la cantidad original.

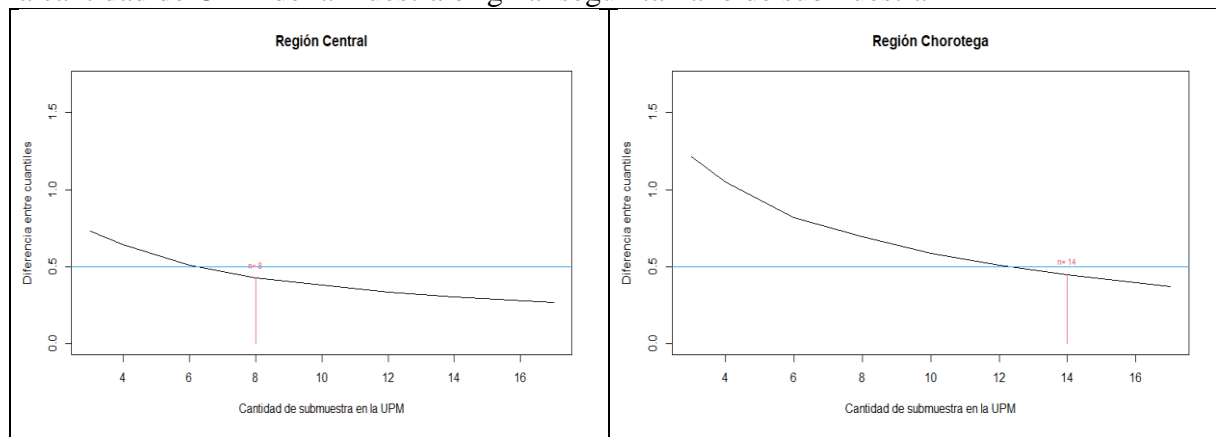
Figura 8. Comportamiento de las medias del índice de uso de Internet según cantidad de submuestra simulada en cada región de planificación

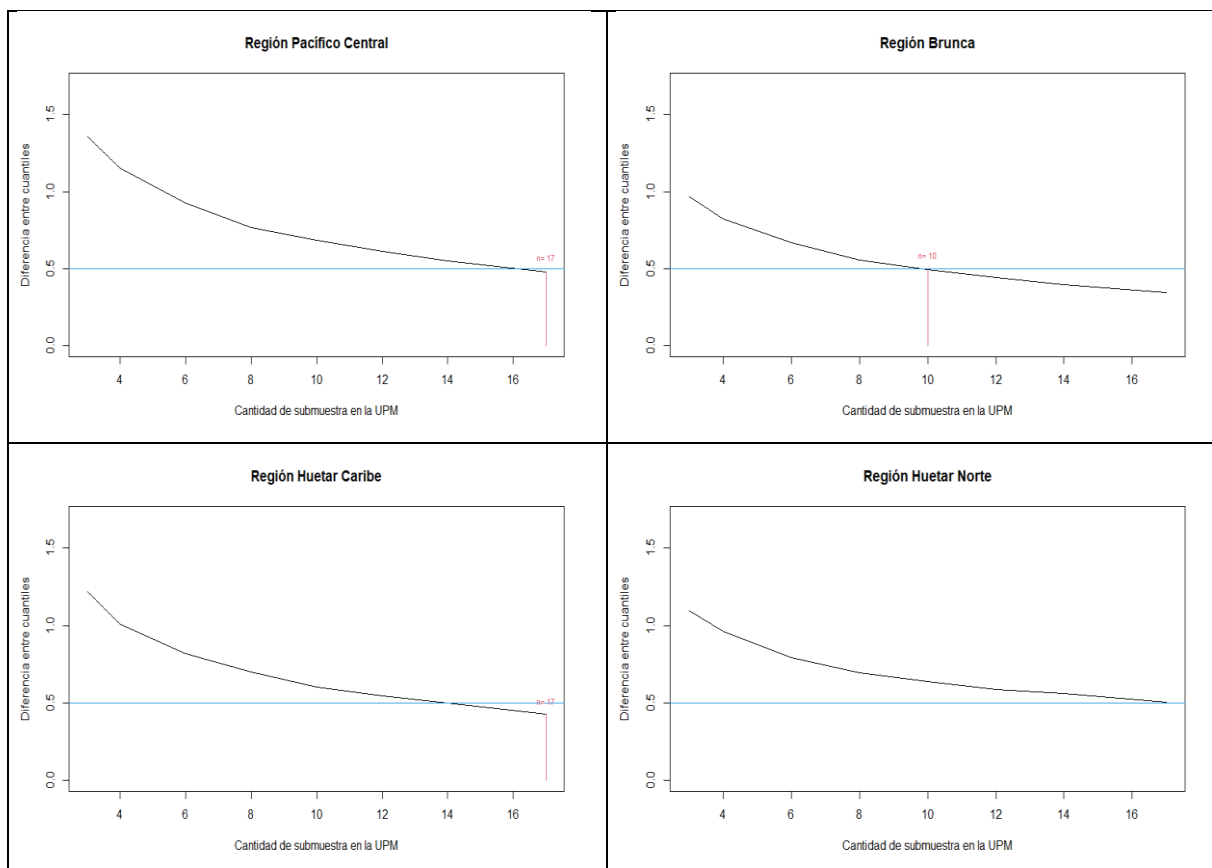


Fuente: MICITT. Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017.

Como complemento a los resultados anteriores, en la Figura 9 se tiene, para cada tamaño de submuestra, la diferencia entre los cuantiles de las medias simuladas del índice de uso de Internet. En la región Central el análisis evidencia que la submuestra se puede disminuir a 8 entrevistas por UPM, reducción de poco más de 50% en la cantidad de entrevistas. En la región Chorotega el criterio establecido se cumple para una submuestra de tamaño 14. Para la región Pacífico Central la submuestra se establece en 17 entrevistas, lo que equivale a una pequeña disminución con respecto a las 19 entrevistas que en promedio se realizan en cada UPM de esta región. En la región Brunca la reducción es importante, pues la submuestra se establece en 10 entrevistas. Para la región Huetar Caribe, el criterio se cumple a partir de la submuestra de tamaño 17. Finalmente, en la región Huetar Norte, se aprecia que la línea prácticamente está pegando con el límite de la precisión deseada, por lo que se considera un tamaño de submuestra aceptado de 17 entrevistas.

Figura 9. Diferencia entre los cuantiles de las medias del índice de uso de Internet utilizando la cantidad de UPM de la muestra original según tamaño de submuestra





Fuente: MICITT. Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017.

En el Cuadro 17 se comparan los resultados del tamaño de muestra, considerando la reducción del número de UPM del proceso anterior, esto es manteniendo la submuestra por UPM con el tamaño original. Estos resultados se comparan con los obtenidos al variar el tamaño de la submuestra y manteniendo la cantidad de UPM de la muestra original.

La columna (1) presenta la cantidad de UPM de la muestra original y la columna (3) el tamaño de muestra original. La columna (2) se obtiene de dividir la columna (3) entre la columna (1). La columna (4) presenta la cantidad de UPM óptima que se obtiene del proceso de simulación y en la columna (5) se tiene el tamaño de muestra con esa cantidad de UPM y el submuestreo original (sale de la multiplicación de la columna (2) y (4)). Finalmente, en la columna (6) se tiene la submuestra simulada utilizando la cantidad de UPM de la muestra original y en la columna (7) el tamaño de muestra utilizando esa cantidad de submuestra (se obtiene de multiplicar la columna (1) y (6)).

Obsérvese que después de conocer los resultados del proceso de simulación, la submuestra utilizando la cantidad de UPM de la muestra original, disminuye considerablemente para la región Central, Chorotega y Brunca, donde en promedio la submuestra era de 17, 18 y 17 respectivamente, y pasa a 8, 14 y 10. Si se compara este resultado con el procedimiento alternativo de optar por un número óptimo de UPM y mantener el submuestreo por UPM originalmente establecido, se observa que la diferencia es 68 entrevistas.

Cuadro 17. Distribución de la muestra original, con UPM simuladas y con submuestra simuladas, según región de planificación, utilizando las estimaciones del índice de uso de Internet

Región de planificación	Cantidad de UPM de la muestra (1)	Tamaño promedio original de la submuestra por UPM (2)	Tamaño de muestra (3)	Cantidad de UPM obtenidas en la simulación (4)	Tamaño de muestra con UPM simuladas (5)	Submuestra utilizando cantidad de UPM de la muestra (6)	Tamaño de muestra con submuestra simulada (7)
Central	60	17	1 006	24	402	8	480
Chorotega	28	18	492	18	316	14	392
Pacífico Central	27	19	520	27	520	17	459
Brunca	29	17	495	19	324	10	290
Huetar Caribe	27	19	508	21	395	17	459
Huetar Norte	32	19	598	32	598	17	544
Total	203		3 618	141	2 556		2 624

Fuente: MICITT. Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017.

De ahí que, en términos de reducción de la muestra, se considera que resulta más factible optar por la simulación que provee una cantidad óptima de UPM (columna 4) y mantener el tamaño promedio de la submuestra por UPM (columna 2), ya que, para efectos logísticos y económicos, aunque la submuestra sea mayor, se concentra en una misma zona geográfica, lo que disminuye los gastos que implican el traslado, transporte y por ende hay una reducción de tiempo, posiblemente de personal humano, y todo esto disminuye el costo final del trabajo de campo.

VI. CONCLUSIONES

La evaluación del tamaño de muestra de la EAU, utilizando un proceso de simulación, brinda una herramienta útil para el planteamiento de la encuesta en años futuros. Como variables de interés para la EAU se estableció el índice de uso de internet y 5 escalas relacionadas con el

uso de Internet. Para cada región de planificación y para las variables antes mencionadas, se realizaron 228 procesos de simulación con 5 000 réplicas cada uno. Esto permitió calcular, para cada muestra generada, la media de las variables, por lo que para valorar la precisión se calcularon los percentiles 2,5 y 97,5 de las 5 000 repeticiones y se estableció que la diferencia entre el máximo y el mínimo no debía sobrepasar la cantidad de 0,5.

Para analizar el número óptimo de UPM en la muestra se calculó el número requerido en cada región, dejando sin cambios el número de entrevistas por UPM realizado en la encuesta. En este procedimiento el número de UPM en la muestra podía aumentar o disminuir. Los resultados evidencian la posibilidad de una reducción importante en la cantidad de UPM, al priorizar las estimaciones del índice de uso de Internet. Para la región Central puede lograrse una reducción del 60%, pasando de 60 a 24 UPM, lo que representa 603 entrevistas menos en el tamaño final de la muestra. Por otro lado, la cantidad de UPM utilizada en las regiones Chorotega, Brunca y Huetar Caribe puede reducirse considerablemente en 36%, 34% y 22% respectivamente, lo que representa una reducción en las tres regiones de 462 entrevistas. En las regiones Pacífico Central y Huetar Norte no se alcanza el criterio establecido, sin embargo, puede verificarse gráficamente que puede mantenerse la misma cantidad de UPM en la muestra sin perder precisión de manera considerable.

El análisis del tamaño de la submuestra se realiza utilizando únicamente las estimaciones del índice de uso de Internet y se analizaron dos posibilidades. La primera consistió en utilizar el número de UPM óptimo obtenido anteriormente para determinar el número de entrevistas requeridas por UPM. La segunda consideró mantener fijo el número de UPM utilizadas originalmente en la muestra y calcular el número de entrevistas requeridas por UPM.

En el primer caso, al analizar el tamaño de la submuestra con la reducción en la cantidad de UPM se obtuvo que no fue posible disminuir la cantidad de entrevistas por UPM, debido a que el número actualmente utilizado es el que alcanza la precisión deseada por región.

En el segundo caso, es decir, al analizar la submuestra utilizando la cantidad de UPM de la muestra original se obtiene que, la muestra final no se reduce. Por lo anterior, termina siendo

más factible, en términos logísticos y económicos, disminuir la cantidad de UPM y mantener el tamaño original de la submuestra por UPM. Esto significa, concretamente, un número de UPM y de entrevistas por UPM, como se muestra en el Cuadro 18.

Cuadro 18. Propuesta del nuevo tamaño de muestra para la EAU

Región de planificación	Número de UPM	Entrevistas por UPM	Tamaño de muestra
Central	24	17	402
Chorotega	18	18	316
Pacífico Central	27	19	520
Brunca	19	17	324
Huetar Caribe	21	19	395
Huetar Norte	32	19	598
Total	141		2556

Fuente: MICITT. Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017.

Es importante tomar en cuenta que la planeación de la encuesta implica evaluar múltiples factores. Uno de los de los más importantes es el presupuesto y el tiempo disponible para la ejecución del trabajo de campo, lo que impactará de manera directa el tamaño de la muestra final. Se considera que los resultados de este estudio permitirán continuar profundizando en el análisis de los datos, incluso incluyendo otras variables que sean de interés para el MICITT. Esto permitirá planificar de una manera más eficiente el trabajo de campo, así como lograr una distribución óptima en el uso de los recursos disponibles.

VII. BIBLIOGRAFÍA

Aguilar, V (2021). Caracterización de los usuarios a partir de escalas de uso de Internet. (Informe de práctica profesional no publicado). Universidad de Costa Rica.

Asamblea Legislativa de la República de Costa Rica (2008), Ley N°8642. Ley General de Telecomunicaciones. San José, Costa Rica.

Bekoff, M. y Mech, L. (1984). Simulation analyses of space use: Home range estimates, variability, and sample size. *Behavior Research Methods, Instruments, & Computers*, 16(1), 32-37. <https://link.springer.com/article/10.3758/BF03201046#Bib1>

Bú, R. C. (1993). *Simulación: Un enfoque práctico*. México: Editorial Limusa.

Brenner, M. (1965). A Relation between Decision Making Penalty and Simulation Sample Size for Inventory Systems. *Operations Research*, 13(3), 433-443. <http://www.jstor.org.ezproxy.sibdi.ucr.ac.cr:2048/stable/167807>

Chandra, S., Huaman, Z., Hari Krishna, S. y Ortiz, R. (2002). Optimal sampling strategy and core collection size of Andean tetraploid potato based on isozyme data – a simulation study. *Theor Appl Genet*, 104(1), 1325-1334. Recuperado de <https://link-springer-com.ezproxy.sibdi.ucr.ac.cr/content/pdf/10.1007%2Fs00122-001-0854-4.pdf>

Fishman, G. (1971). Estimating Sample Size in Computing Simulation Experiments. *Management Science*, 18(1), 21-38. Recuperado de <http://www.jstor.org.ezproxy.sibdi.ucr.ac.cr:2048/stable/2629290>

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.

Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement*, 30(3), 607-610

Lenth, R. V. (2001). Some practical guidelines for effective sample size determination. *The American Statistician*, 55(3), 187-193.

Mendell, N., Thode, H. y Finch, S. (1991). The Likelihood Ratio Test for the Two-Component Normal Mixture Problem: Power and Sample Size Analysis. *Biometrics*, 47(3), 1143-1148. Recuperado de <https://www-jstor-org.ezproxy.sibdi.ucr.ac.cr/stable/2532666>

Ministerio de Ciencia, Tecnología y Telecomunicaciones (2019). *Acceso y Uso de los Servicios de Telecomunicaciones en Costa Rica 2017*. San José, Costa Rica.

Robert, C. P., & Casella, G. (2010). *Introducing Monte Carlo methods with R*. Springer Science & Business Media

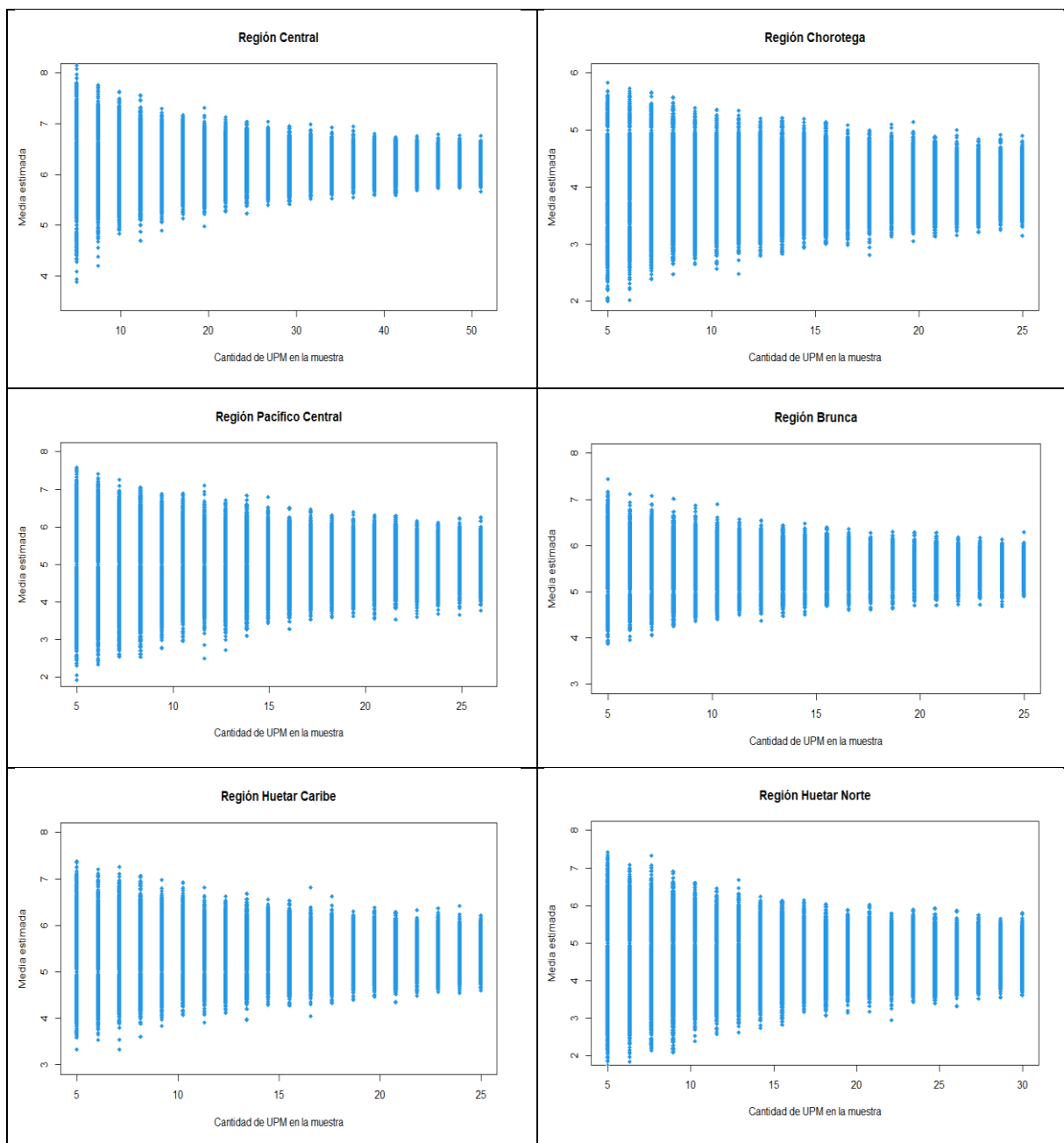
VIII. ANEXOS

Anexo 1. Batería de preguntas utilizadas para la construcción del índice y las escalas de uso de Internet

Descargar imágenes, videos, música, jugar o descargar juegos
Descargar software o aplicaciones
Participar en redes sociales
Descargar aplicaciones
Llamar por teléfono por internet/VoIP
Enviar o recibir correo electrónico (uso privado/personal)
Buscar información sobre salud
Obtener información sobre bienes y servicios
Colgar texto, imágenes, fotos, videos, música en un sitio web para compartirlo
Leer o descargar periódicos o revistas en línea, libros electrónicos
Obtener información de sitios web gubernamentales o de servicios públicos
Escuchar la radio por Internet
Consultar Wikipedia, enciclopedias en línea
Ver televisión por la Internet
Enviar o recibir correo electrónico (trabajo)
Realizar trámites u operaciones bancarias
Utilizar espacio de almacenamiento en Internet para guardar documentos
Buscar empleo o enviar solicitudes de empleo
Descargar formularios oficiales de sitios web gubernamentales o de servicios públicos
Comprar o encargar bienes o servicios
Acceder a sitios de charla, blog, grupo de noticias o debates en línea
Edición en Internet de documentos, hojas de cálculo o presentaciones
Enviar formularios completos de sitios web gubernamentales o de servicios públicos
Utilizar servicios a viajes y el correspondiente alejamiento
Actualizar o añadir contenido a los blogs
Cursar estudios oficiales en línea
Vender bienes o servicios
Cargar contenido creado de forma independiente en sitios web compartidos
Participar en redes de profesionales
Interactuar con organizaciones gubernamentales generales
Publicar opiniones sobre cuestiones políticas o cívicas en los sitios web
Participar en consultas en línea o votar sobre asuntos cívicos o políticos
Gestionar una página web personal/propia

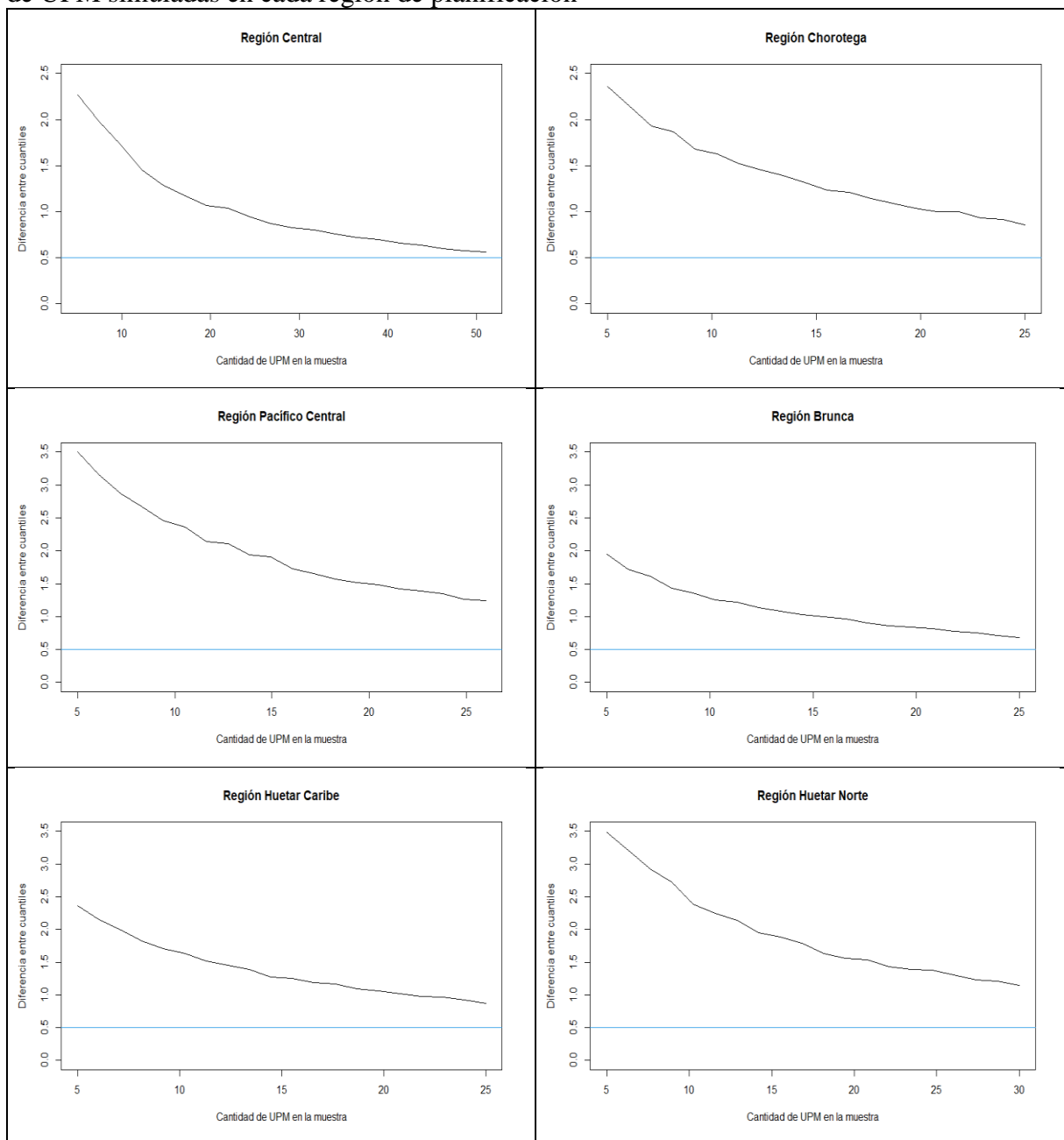
Fuente: MICITT. Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017.

Anexo 2. Comportamiento de las medias de la escala informativa según número de UPM simuladas en cada región de planificación



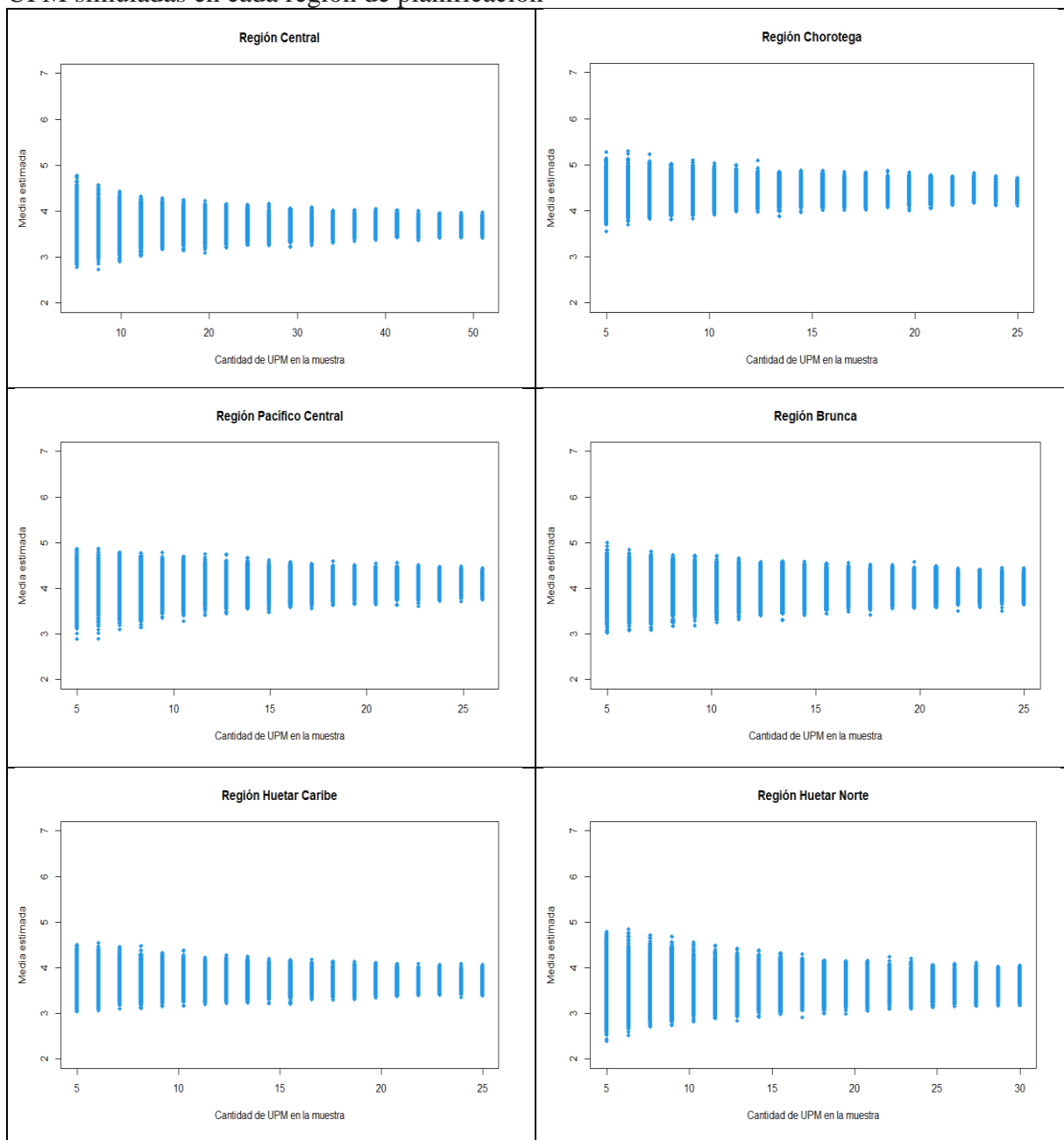
Fuente: MICITT. Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017.

Anexo 3. Diferencia entre los cuantiles de las medias de la escala informativa según cantidad de UPM simuladas en cada región de planificación



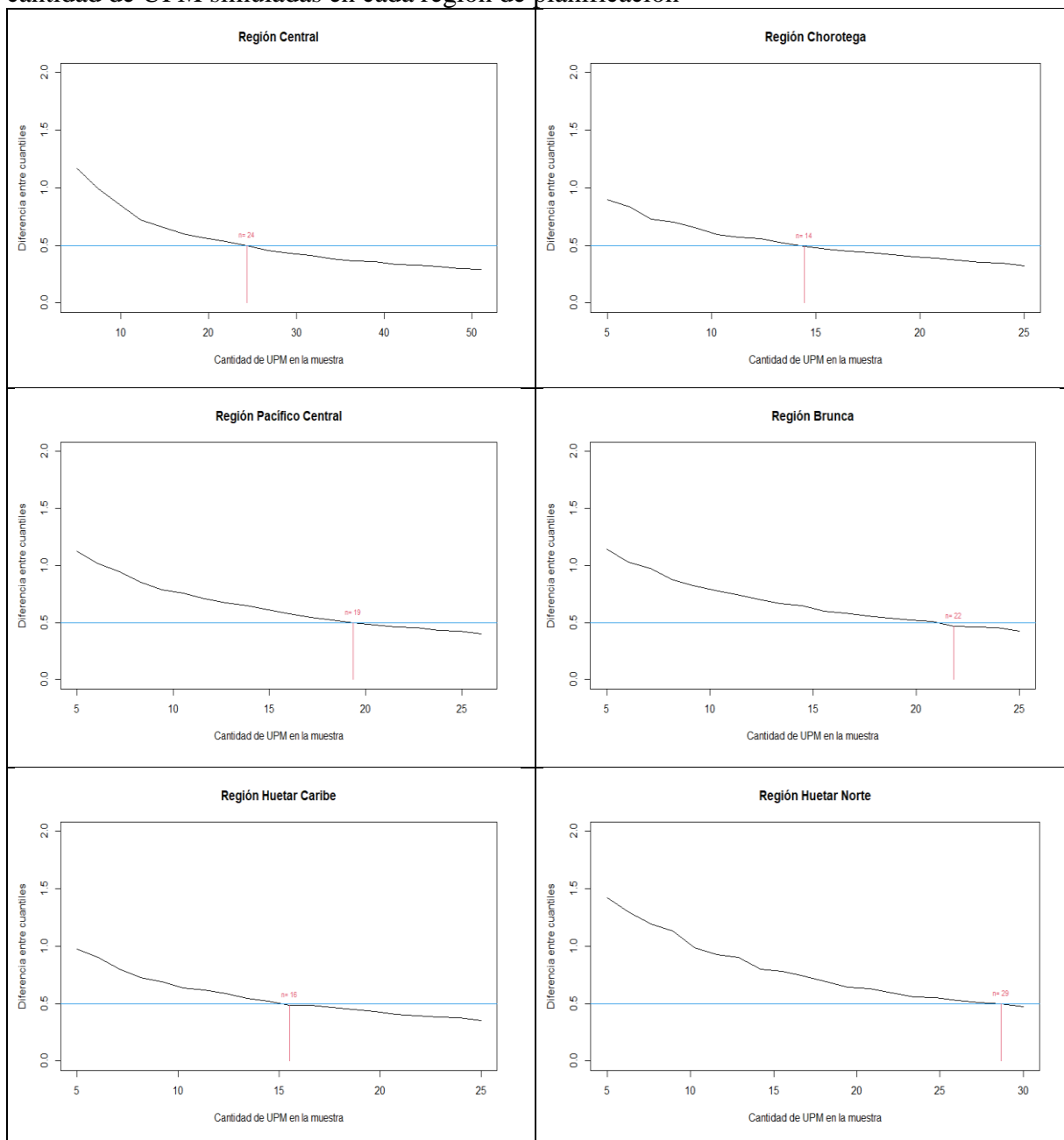
Fuente: MICITT. Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017.

Anexo 4. Comportamiento de las medias de la escala de interacción social según número de UPM simuladas en cada región de planificación



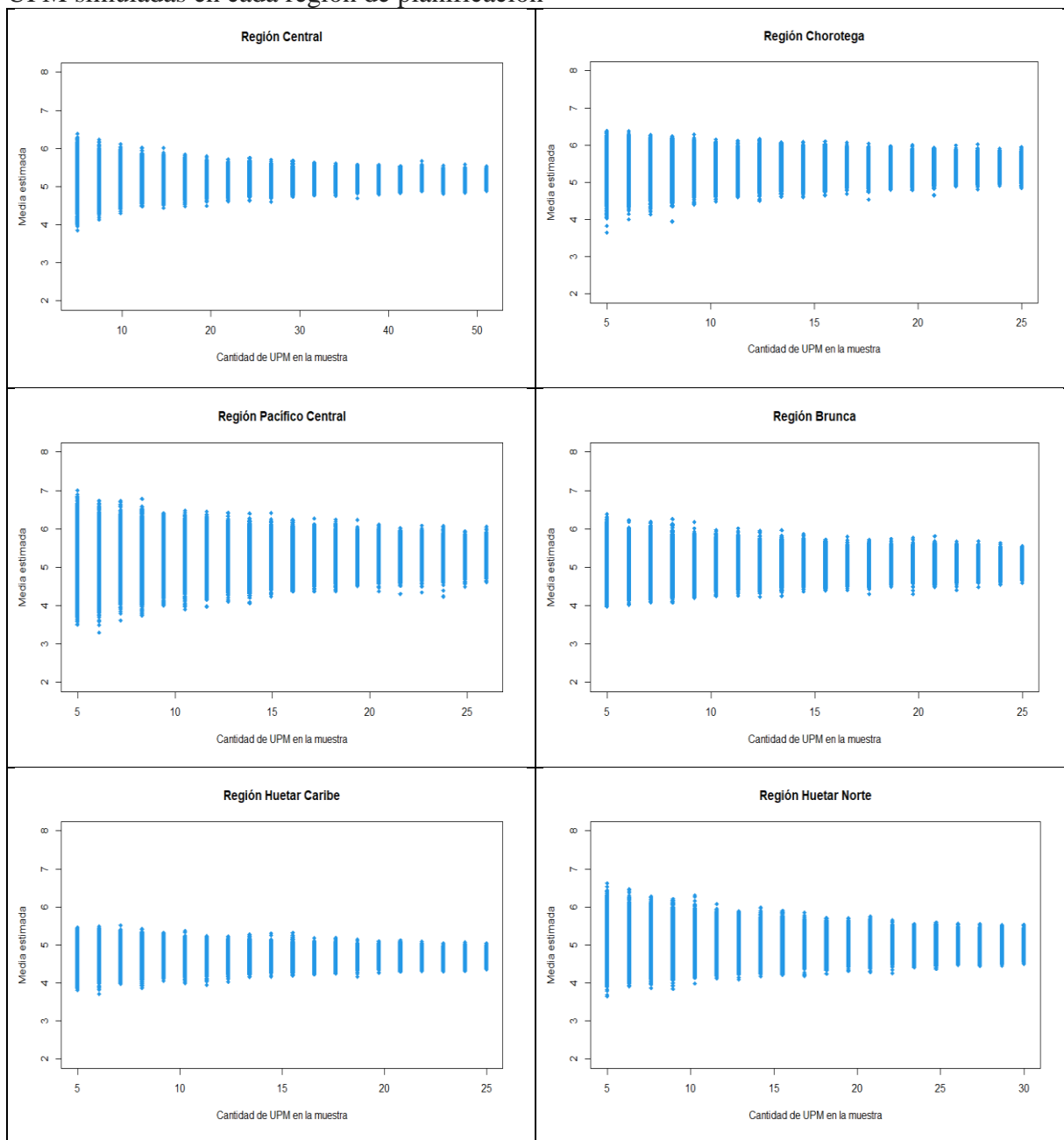
Fuente: MICITT. Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017.

Anexo 5. Diferencia entre los cuantiles de las medias de la escala de interacción social según cantidad de UPM simuladas en cada región de planificación



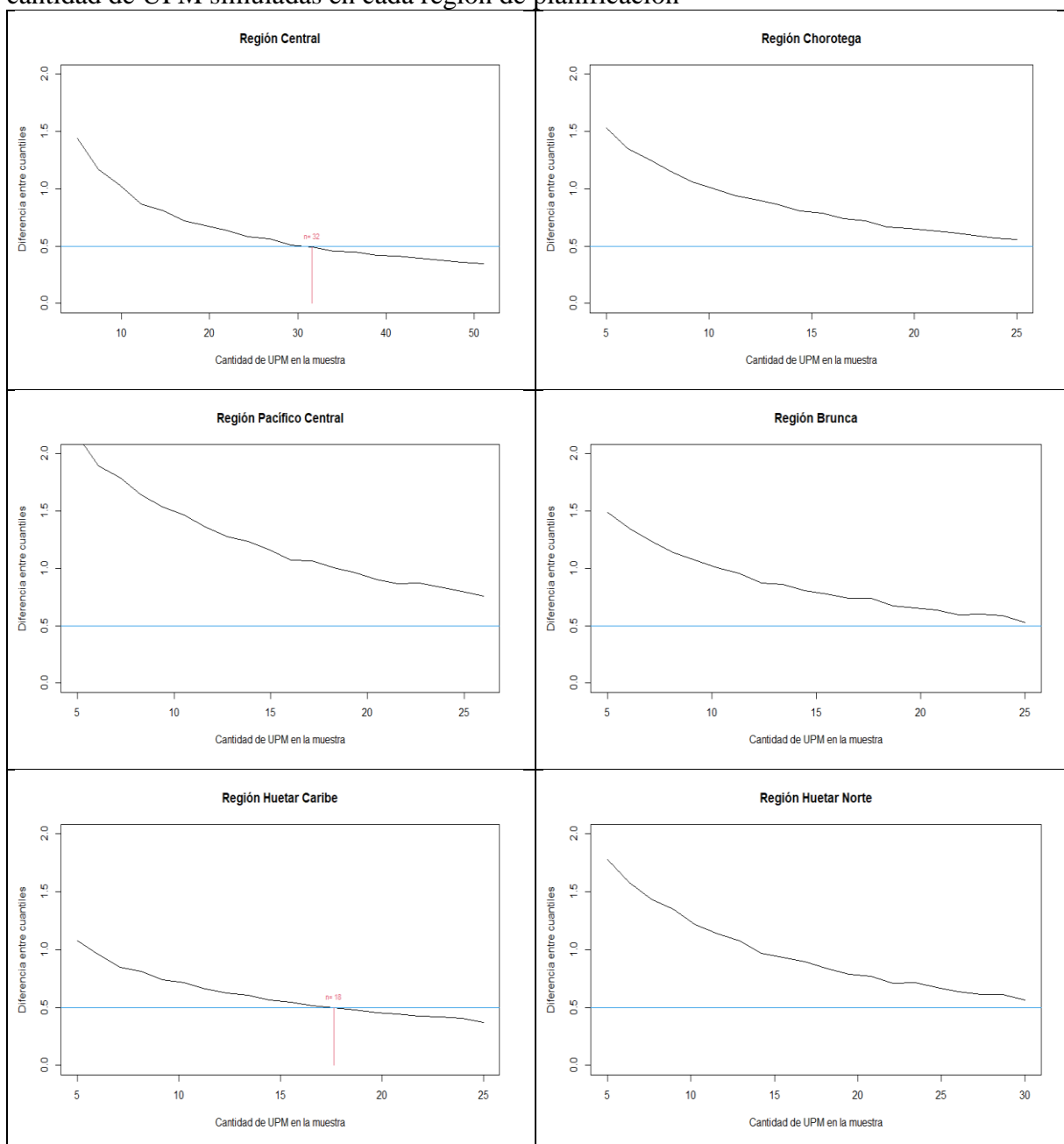
Fuente: MICITT. Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017.

Anexo 6. Comportamiento de las medias de la escala de entretenimiento según número de UPM simuladas en cada región de planificación



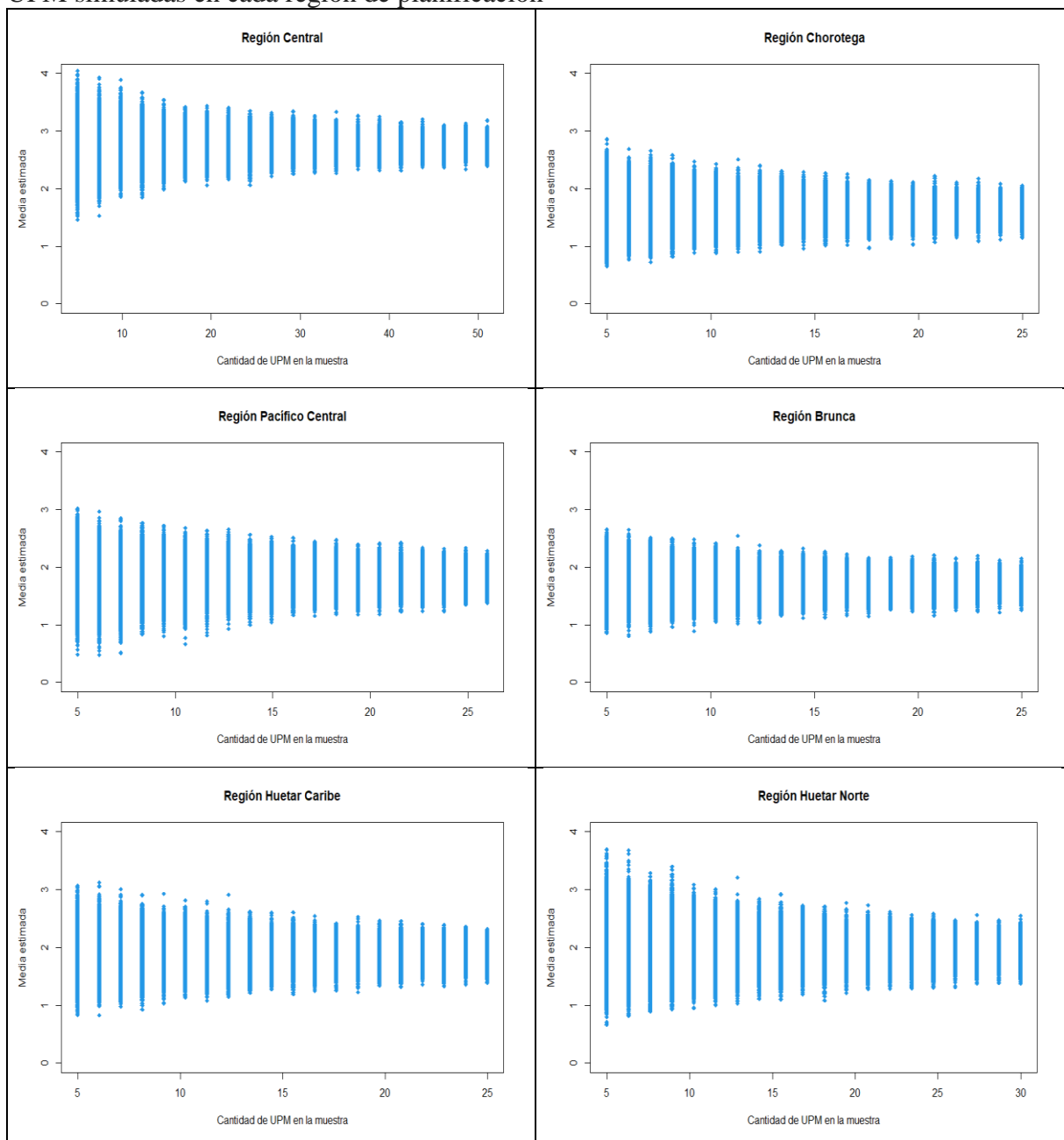
Fuente: MICITT. Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017.

Anexo 7. Diferencia entre los cuantiles de las medias de la escala de entretenimiento según cantidad de UPM simuladas en cada región de planificación



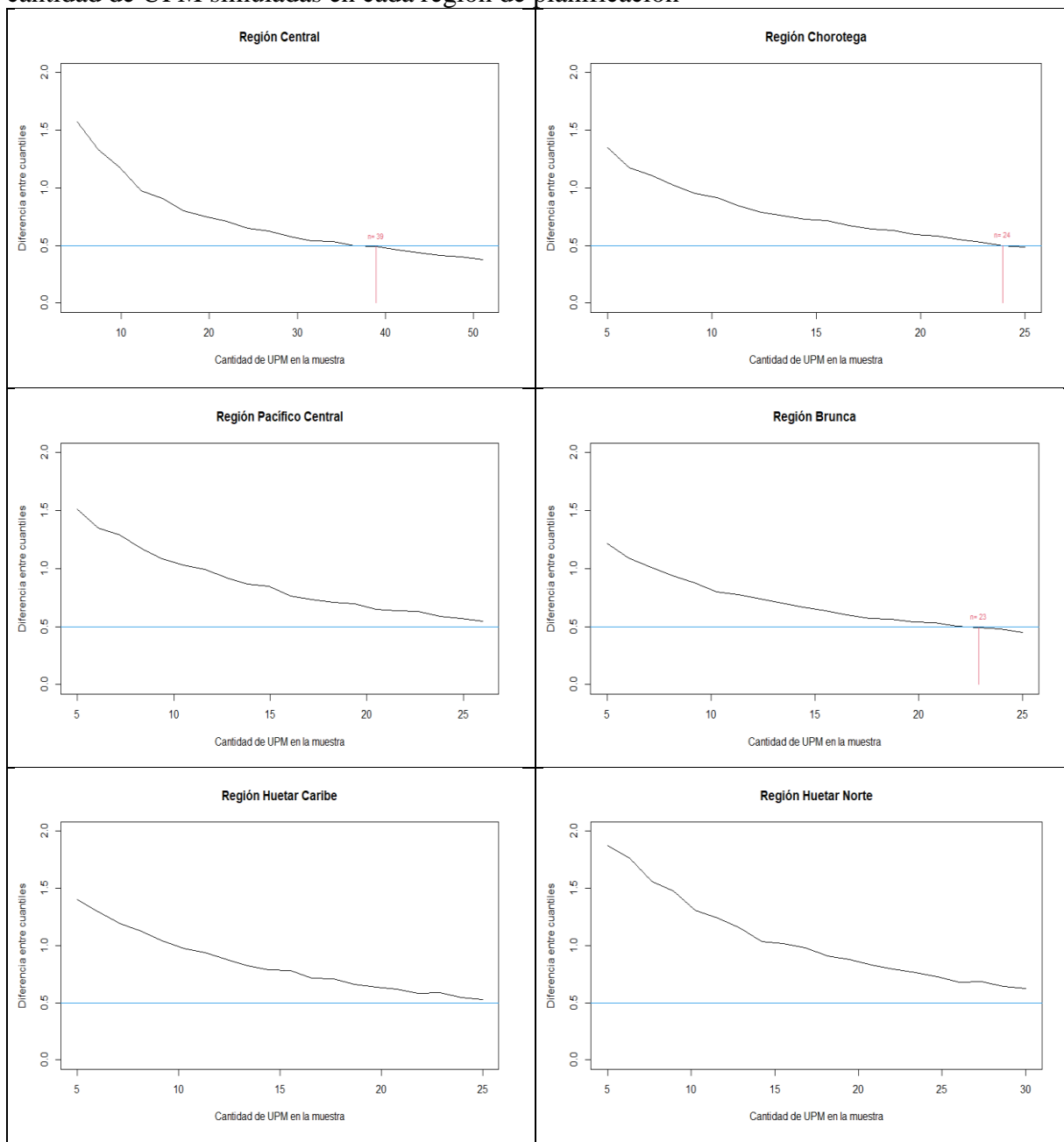
Fuente: MICITT. Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017.

Anexo 8. Comportamiento de las medias de la escala de productividad según número de UPM simuladas en cada región de planificación



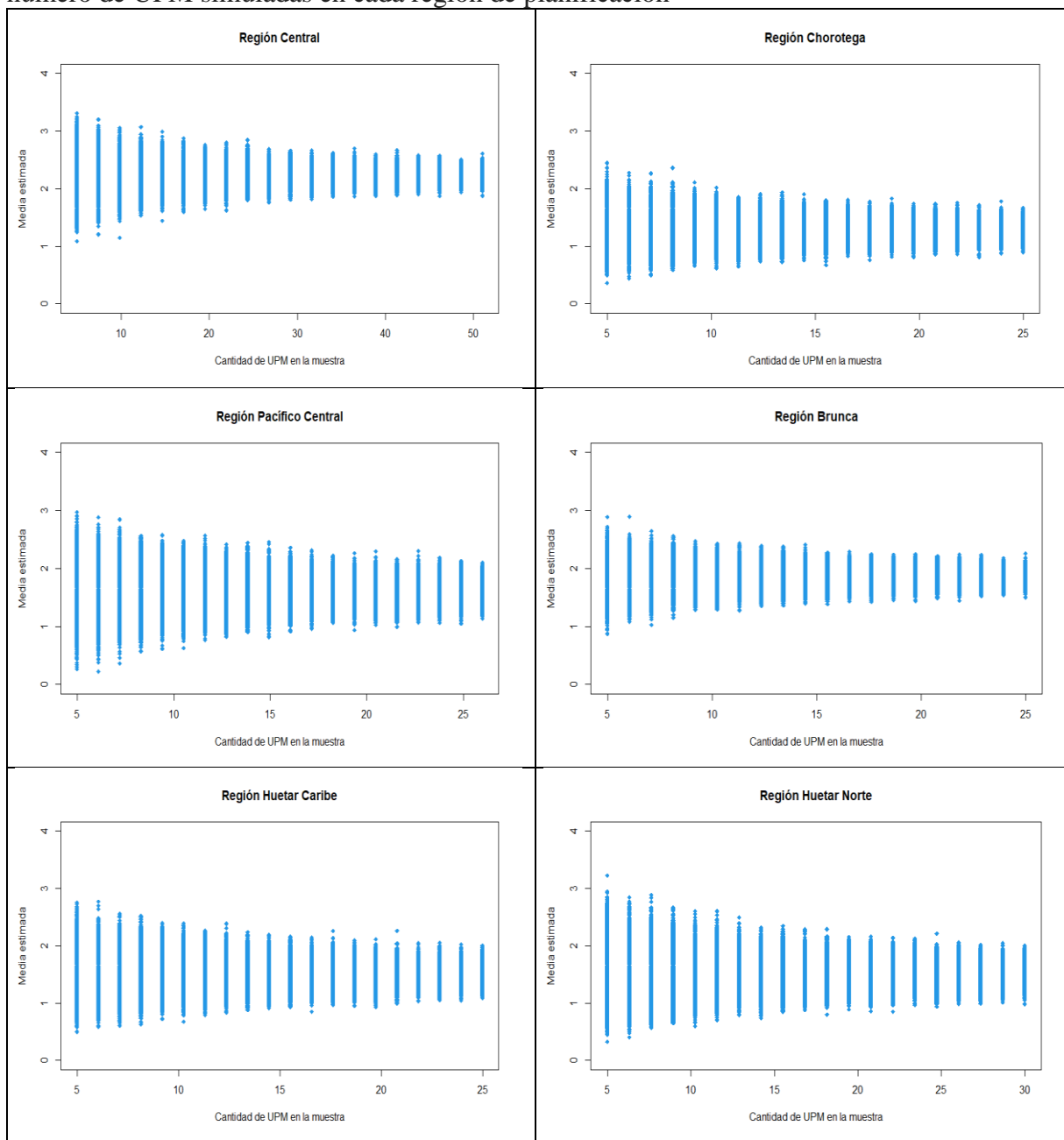
Fuente: MICITT. Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017.

Anexo 9. Diferencia entre los cuantiles de las medias de la escala de productividad según cantidad de UPM simuladas en cada región de planificación



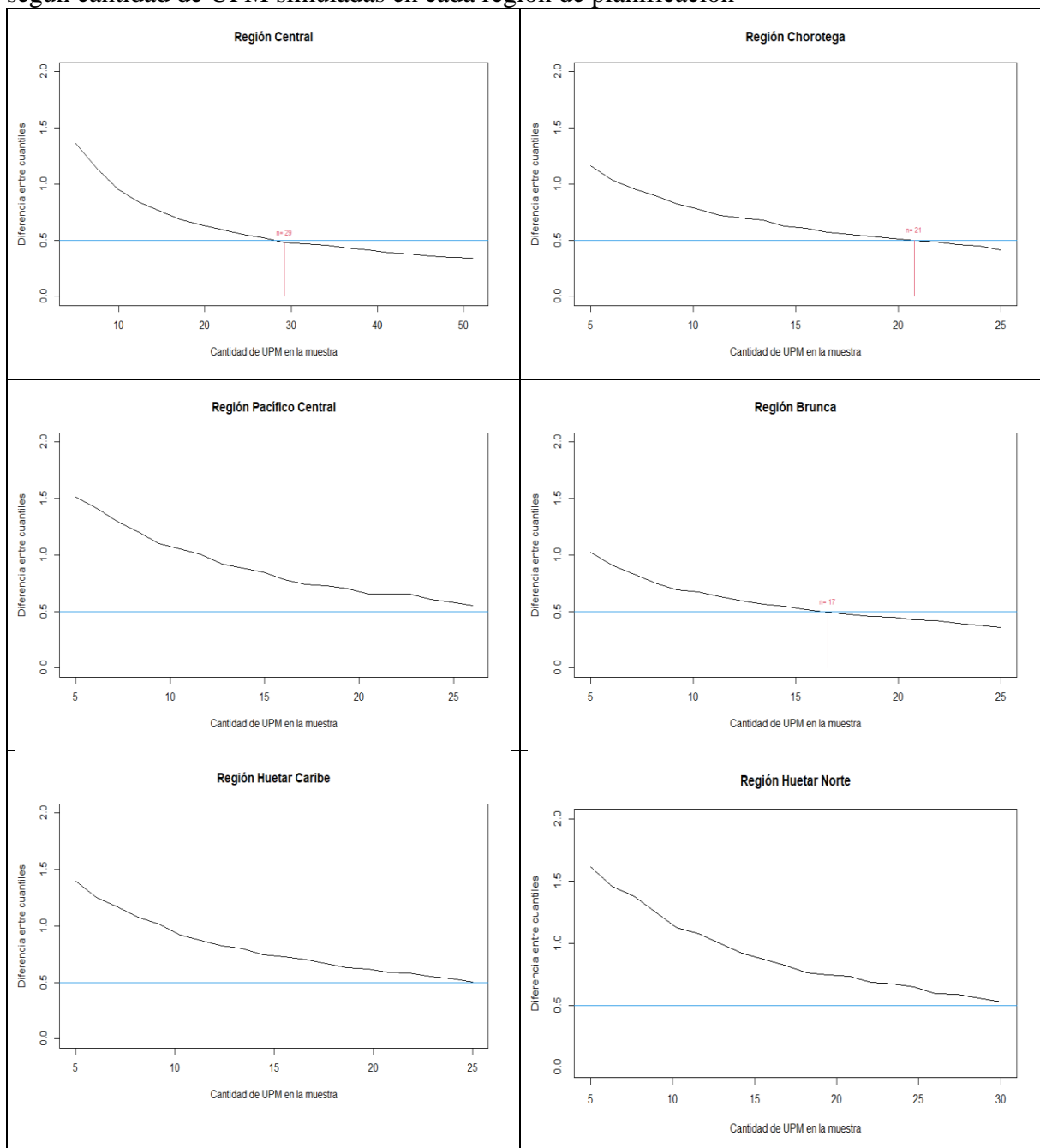
Fuente: MICITT. Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017.

Anexo 10. Comportamiento de las medias de la escala de habilitación ciudadana según número de UPM simuladas en cada región de planificación



Fuente: MICITT. Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017.

Anexo 11. Diferencia entre los cuantiles de las medias de la escala de habilitación ciudadana según cantidad de UPM simuladas en cada región de planificación



Fuente: MICITT. Encuesta de acceso y uso de los servicios de telecomunicaciones en Costa Rica 2017.