



UNIVERSIDAD DE LA REPÚBLICA
Facultad de Ciencias Económicas y de Administración
Licenciatura en Estadística
Informe de Pasantía

Monitor de indicadores socioeconómicos obtenidos a partir de la Encuesta Continua de Hogares

Mauricio Pittamiglio

Tutores:

Ignacio Álvarez-Castro

Juan José Goyeneche

Montevideo, Fecha.

Índice general

Índice general	III
Índice de figuras	v
Índice de tablas	1
1. Introducción	3
2. Indicadores socioeconómicos	5
2.1. Encuesta Continua de Hogares (ECH)	5
2.2. Indicadores	7
3. Marco teórico	9
3.1. Notación y definiciones	9
3.2. Inferencia	11
3.2.1. Estimaciones puntuales	11
3.2.2. Intervalos de confianza	12
3.3. Estimación de los errores estándar (SE)	13
3.3.1. Aproximación de una etapa	14
3.3.2. Linearización de Taylor	15
4. Métodos computacionales	17
4.1. Estimación de los indicadores	17
4.1.1. Descripción de los diseños	21
4.2. Aplicación web utilizando Shiny	21

ÍNDICE GENERAL

4.2.1. Monitor de indicadores socioeconómicos	22
4.2.2. Interfaz del usuario	23
4.2.3. Comportamiento reactivo	26
5. Ejemplo de aplicación	29
5.1. Tasa de desempleo	29
5.2. Hogares por presencia de lugar para cocinar	34
6. Discusión	37
6.1. Limitaciones y trabajos a futuro	38
Bibliografía	41
A. Anexos	43
A.1. Listado de indicadores	43
A.2. Listado de variables	48

Índice de figuras

2.1. Histograma de la cantidad de UPM por estrato.	6
4.1. Barra lateral.	24
4.2. Ventana inicial de la app.	25
4.3. Sección indicadores.	25
5.1. Gráfico de barras de la tasa de desempleo para el año 2019 a nivel departamental y de todo el país desagregando por sexo.	30
5.2. Tasa de desempleo a nivel departamental para la categoría Hombres.	31
5.3. Tabla con las estimaciones puntuales de la tasa de desempleo para los 19 departamentos y para todo el país en el año 2019.	32
5.4. Estimación de la tasa de desempleo en hombres para los 19 departamentos y para todo el país en el año 2019, con los respectivos intervalos de confianza al 95 %.	33
5.5. Tabla con la estimación de la tasa de desempleo en hombres para los 19 departamentos y para todo el país en el año 2019, con los respectivos intervalos de confianza al 95 %.	33
5.6. Gráfico de barras apiladas al 100 % de la proporción de hogares por presencia de lugar para cocinar por departamento para el año 2019.	35
5.7. Hogares con lugar privado para cocinar para los 19 departamentos y para todo el país en el año 2019, con los respectivos intervalos de confianza al 95 %.	36

ÍNDICE DE FIGURAS

Índice de tablas

A.1. Listado de indicadores	43
A.2. Listado de variables utilizadas en la app	48

ÍNDICE DE TABLAS

Capítulo 1

Introducción

Las encuestas por muestreo son utilizadas no solo para obtener estimaciones a nivel de toda la población, sino que también pueden ser eficaces para realizar estimaciones en subconjuntos de la población. Estos subconjuntos son denominados dominios, y pueden estar definidos por áreas geográficas, grupos demográficos u otro tipo de subpoblaciones.

En Uruguay, la principal fuente de información para construir indicadores socio-demográficos lo constituye la Encuesta Continua de Hogares (ECH) realizada anualmente por el Instituto Nacional de Estadística (INE). Dicha encuesta cuenta con un diseño muestral en base al cual el INE presenta estimaciones de algunos indicadores relevantes a nivel nacional, como por ejemplo la tasa de desempleo. La información de microdatos producida por la ECH es de acceso abierto, por lo que cualquier investigador puede obtener esos datos y utilizarlos directamente. Sin embargo, en el caso de usuarios no-expertos en el manejo de datos, únicamente acceden a los indicadores relevados directamente por el INE.

El objetivo de este trabajo es generar un **Monitor de indicadores socio-económicos**, una herramienta computacional que facilite el procesamiento de la ECH mediante el cálculo de diversos indicadores y sus respectivos intervalos de confianza, aportando además una visualización interactiva de los resultados. En primer lugar, se busca extender la cantidad de indicadores presentados en el informe del INE,

CAPÍTULO 1. INTRODUCCIÓN

incluyendo además desagregaciones departamentales en todos ellos. Se trabajará con distintos indicadores socioeconómicos que se obtienen a partir de estimaciones provenientes de la Encuesta Continua de Hogares. Estos indicadores abordan distintas dimensiones, como por ejemplo, educación, salud y mercado laboral, entre otras. Un segundo aspecto relevante del monitor es su disponibilidad en la web, haciendo posible para un usuario final acceder a los resultados (estimaciones, intervalos, gráficos) sin necesidad de programar directamente. El monitor se encuentra disponible en el siguiente repositorio en GitHub. Además está disponible directamente para su uso, una versión reducida del mismo, en el siguiente link APP. Esta versión reducida permite el cálculo de una menor cantidad de indicadores que la versión original.

Este trabajo se estructura de la siguiente manera. En el capítulo 2 se describe el diseño de la ECH y se presentan los indicadores trabajados. En el capítulo 3 se exponen los métodos estadísticos utilizados en este trabajo, los conceptos y las fórmulas que se aplican tanto para las estimaciones puntuales como para los errores estándar e intervalos de confianza. El capítulo 4 refiere a la computación estadística, explicando la forma de cálculo de los indicadores, las características y el funcionamiento de la aplicación. En el capítulo 5 se presenta un ejemplo de aplicación de la app en dos indicadores. Finalmente en el capítulo 6 se realiza la discusión final, presentando los posibles trabajos a futuro.

Buena parte de los resultados presentados en este trabajo fueron desarrollados en el marco del convenio entre IESTA y la Oficina de Planeamiento y Presupuesto (OPP), donde trabajé como integrante del equipo durante los años 2019 y 2020. Mis tareas consistieron principalmente en el cálculo de indicadores a partir de la ECH, en este contexto fue notorio que contar con una herramienta como el monitor que se presenta en este trabajo simplifica considerablemente la tarea. Parte de los resultados de este convenio fueron presentados en el siguiente informe (Coimbra, A., Goyeneche, J.J., Ferreira, J.P., Pittamiglio, M., 2020).

Capítulo 2

Indicadores socioeconómicos

2.1. Encuesta Continua de Hogares (ECH)

La ECH es una encuesta realizada de forma ininterrumpida por el Instituto Nacional de Estadística (INE) desde el año 1968. La ECH es de interés nacional, ya que brinda los indicadores oficiales del mercado laboral y de ingresos de los hogares y las personas con periodicidad mensual, trimestral, semestral y anual, entre otros.

Esta encuesta cuenta con un diseño muestral aleatorio complejo, donde se incluyen estratos, conglomerados y se cuenta con varias etapas de selección. A continuación se realiza una breve descripción del mismo, la descripción completa se puede encontrar en INE. (2020), disponible en la Ficha técnica Encuesta Continua de Hogares.

En primer lugar se divide a la población en estratos geográficos que corresponden a los diecinueve departamentos y a la zona metropolitana. En un segundo nivel de estratificación, se clasifica a todas las localidades dentro de cada departamento en cuatro categorías según la cantidad de habitantes.

- Localidades de 20000 o más habitantes
- Localidades de entre 5000 y 20000 habitantes
- Localidades de entre 200 y 5000 habitantes

CAPÍTULO 2. INDICADORES SOCIOECONÓMICOS

- Áreas rurales y localidades con menos de 200 habitantes.

Además en los departamentos de Maldonado y Rocha se agregan estratos correspondientes a las zonas balnearias. Montevideo, por su parte, queda particionado en cinco estratos socioeconómicos, mientras la zona metropolitana se divide en tres.

Una vez definidos los estratos, se selecciona de manera independiente dentro de cada uno de ellos, conglomerados de viviendas (zonas censales), las cuales se denominan unidades primarias de muestreo (UPM). Estas se seleccionan bajo un muestreo sistemático y proporcional al tamaño (PPS), utilizando como medida de tamaño (MDT) la cantidad de viviendas según el censo de población y vivienda del año 2011. A continuación se presenta un gráfico donde se observa la distribución de la cantidad de UPM_s dentro de cada estrato.

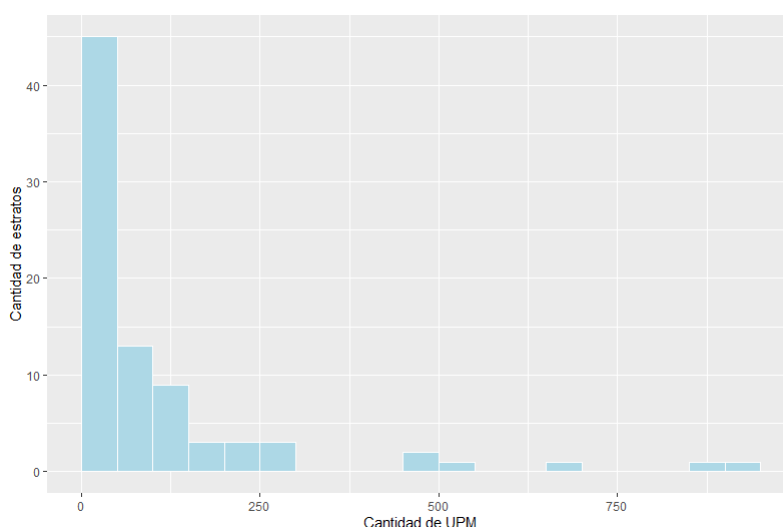


Figura 2.1: Histograma de la cantidad de UPM por estrato.

El ancho de los bins es de 50, por lo que podemos observar que más de la mitad de los estratos (45 de los 82) tienen menos de 50 UPM_s . Los estratos que contienen mayor cantidad de UPM_s son los pertenecientes a Montevideo y al área metropolitana.

En la segunda etapa, dentro de cada UPM seleccionada en la primera etapa, se seleccionan cinco viviendas titulares y dos viviendas suplentes bajo un muestreo

aleatorio simple.

Dentro de cada vivienda seleccionada se relevan características propias de la vivienda así como información de todos los integrantes que conforman el hogar. Esto implica que la ECH brinde información para calcular distintos indicadores no solo a nivel de hogar sino también a nivel de persona. Es por esto, que el INE publica dos bases de datos, una base con información a nivel del hogar y vivienda y otra base con información a nivel de personas.

El tamaño de muestra efectivo anual de la ECH, es decir, la cantidad de hogares que fueron seleccionados en la muestra y respondieron a la encuesta, se encuentra en el entorno de 42000 hogares.

Los datos de la muestra son ponderados con el objetivo de poder realizar estimaciones de distintos indicadores, tanto a nivel nacional, departamental y en otros dominios como el sexo y determinados tramos etarios. En primer lugar se definen los ponderadores originales como el inverso de la tasa de muestreo en el estrato, pero luego son ajustados en base a distintos aspectos, como lo es la no respuesta.

2.2. Indicadores

Los indicadores presentados en este trabajo son calculados todos los años por el Observatorio Territorio Uruguay (OTU) perteneciente a la Oficina de Planeamiento y Presupuesto (OPP), y se encuentran disponibles en su página oficial (OTU). Estos indicadores se dividen en 7 categorías según las dimensiones a las que refieren (algunos indicadores refieren a más de una dimensión):

- Educación
- Salud
- Mercado laboral
- Ingresos y bienestar

CAPÍTULO 2. INDICADORES SOCIOECONÓMICOS

- Tecnología y comunicación
- Demografía
- Viviendas y hogares

Dentro de la categoría *Educación* se encuentran 17 indicadores, referidos a tasas de asistencia a distintos niveles educativos, tasas de analfabetismo, promedio de años de educación, entre otros.

En la categoría *Salud* hay 3 indicadores que evalúan la afiliación a emergencias móviles y el tipo de atención en salud.

Hay 13 indicadores disponibles dentro de *Mercado laboral*, además de los principales indicadores de este rubro (tasas de empleo, desempleo, actividad), se encuentran otros referidos a la informalidad, categoría de ocupación, entre otros.

La categoría *Ingresos y bienestar* cuenta con 8 indicadores, índices de pobreza (en personas y hogares) y algunos indicadores presentes también en la categoría *Mercado laboral* (empleo, desempleo e informalidad).

En *Tecnología y comunicación* se encuentran 7 indicadores referidos a tenencia de distintos dispositivos como celulares o computadoras, la utilización de estos dispositivos y el acceso a internet.

Demografía contiene 3 indicadores referidos a residencias previas o lugar de nacimiento de las personas.

Finalmente la categoría *Viviendas y hogares* cuenta con 8 indicadores, todos ellos referidos a distintas características de los hogares.

Recordando que hay indicadores que se encuentran en varias categorías, en total se trabaja con 53 indicadores socio-económicos, desagregando todos por departamento y a su vez, algunos de estos se desagregan también por sexo o tramos etarios.

En el anexo A.1 se presenta un listado completo de los indicadores trabajados.

Capítulo 3

Marco teórico

En este capítulo se presenta el marco teórico para la inferencia sobre los indicadores presentados en el capítulo anterior. La presentación de los métodos se apoya en (Ferreira, 2020) y el informe (Coimbra, A., Goyeneche, J.J., Ferreira, J.P., Pittamiglio, M., 2020) mencionado anteriormente. En primer lugar se plantea la notación y las principales definiciones aplicadas, luego se presenta la inferencia estadística, es decir las fórmulas aplicadas para la estimación puntual de los indicadores y sus respectivos intervalos de confianza. Finalmente se presentan los métodos utilizados para el cálculo de los errores estándar (SE).

3.1. Notación y definiciones

El objetivo principal de una encuesta por muestreo es obtener estimaciones de distintos parámetros de interés de la población sin observar toda la población, sino un subconjunto de esta, denominado muestra. Es decir, dada la población objetivo U de tamaño N , se selecciona una muestra s de tamaño n . En el caso particular de la ECH, recordando que esta cuenta con una base de hogares y otra de personas, en este último caso, U es la población de Uruguay y N es aproximadamente 3.45 millones.

Por ejemplo, un parámetro de interés puede ser el total de la variable de interés

y , el cual se define como:

$$t = \sum_{i \in U} y_i$$

Donde y_i es el valor que toma la variable y en el individuo i .

La mayoría de los indicadores calculados en este trabajo son tasas o proporciones, las cuales se obtienen a partir del cociente (*ratio*) entre totales de dos variables y y z , quedando definido de la siguiente manera:

$$R = \frac{\sum_{i \in U} y_i}{\sum_{i \in U} z_i}$$

Por otra parte, en la práctica las encuestas nacionales como la ECH, no son solamente utilizadas para hacer estimaciones a nivel de toda la población sino que también se realizan para diferentes subpoblaciones. Estas subpoblaciones, para las cuales se computan de manera específica, tanto estimaciones puntuales como intervalos de confianza, son denominadas dominios (Särndal et al., 2003).

Según (Gutiérrez, 2016) un dominio U_d debe cumplir con dos condiciones fundamentales:

- $U_d \subset U$, tal que $U = \bigcup_{d=1}^D U_d$
- Si $k \in U_l$, entonces $k \notin U_d$ para $d \neq l$

Los dominios pueden clasificarse como planeados o no planeados, según si fueron tenidos en cuenta o no en la planificación de la muestra. Los más comunes en la práctica son los no planeados. Por ejemplo, en el indicador *Adolescentes de 12 a 17 años que asisten a establecimientos educativos*, un dominio de estimación son las personas que tienen entre 12 y 17 años dentro de determinado departamento, siendo un ejemplo de dominio no planeado.

3.2. Inferencia

3.2.1. Estimaciones puntuales

Tanto t como R , definidos anteriormente, son desconocidos, por lo que serán estimados a partir de la muestra. Para ello se define el estimador de Horvitz-Thompson (Horvitz and Thompson, 1952).

$$\hat{t}_\pi = \sum_s w_i \times y_i \quad (3.1)$$

Donde w_i es el peso de la unidad i incluida en la muestra, es decir, w_i indica cuántas unidades de la población U que no fueron seleccionadas, están representadas por la unidad i que sí pertenece a la muestra s . En la base de datos de personas de la ECH 2019, los ponderadores anuales toman valores entre 7 y 57. Este estimador es insesgado para el total $t = \sum_U y_i$ (Horvitz and Thompson, 1952).

La construcción de los ponderadores parten de la probabilidad de inclusión π_i de la unidad i en la muestra s . Más precisamente el ponderador o peso muestral w_i se define como el inverso de la probabilidad de inclusión.

La ECH en particular cuenta con un diseño muestral por conglomerados en dos etapas de selección, pero además es estratificado, es decir que la población U en primera instancia se divide en H estratos y se selecciona una muestra s_h de forma independiente dentro de cada uno de ellos.

Dentro de cada estrato se encuentran M_h UPM_s de las cuales se seleccionan m_h con probabilidad proporcional al tamaño, en base a una medida de tamaño (MDT), que en este caso, surge de la cantidad de viviendas N_j de la UPM según los datos del censo de población del año 2011.

En una segunda etapa, se seleccionan n_{jh} unidades dentro de cada UPM seleccionada previamente, bajo un muestreo aleatorio simple. Cabe aclarar que el número de unidades seleccionadas en cada UPM es fijo, es decir $n_j = \alpha$. Por lo tanto la probabilidad de inclusión en la muestra de una unidad i queda definida como:

$$\pi_{ijh} = \pi_{jh} \times \pi_{i|jh}$$

Donde π_{jh} es la probabilidad de selección de la UPM j en el estrato h definida como:

$$\pi_{jh} = m_h \times \frac{\text{MDT}_j}{\sum_{j=1}^{M_h} \text{MDT}_j}$$

Mientras que $\pi_{i|jh}$ es la probabilidad de selección de la unidad i dentro de la UPM j :

$$\pi_{i|jh} = \frac{n_{jh}}{\text{MDT}_j}$$

Como se mencionó anteriormente, los indicadores trabajados son en su mayoría ratios, por lo que el estimador queda definido de la siguiente manera:

$$\hat{R} = \frac{\sum_{i \in s} w_i \times y_i}{\sum_{i \in s} w_i \times z_i} \quad (3.2)$$

Este estimador no es lineal y es aproximadamente insesgado, ya que es un cociente entre estimadores insesgados.

En toda encuesta suelen existir problemas de no respuesta, esto implica que en la práctica no se utilicen de forma directa los ponderadores para realizar las expansiones, sino que estos son ajustados por no respuesta y posteriormente calibrados a información conocida de la población. Estos últimos ponderadores son los que se encuentran en las bases de datos públicas de la ECH, por lo que serán los utilizados a la hora de realizar las estimaciones.

3.2.2. Intervalos de confianza

Los intervalos de confianza son una medida de calidad de las estimaciones y son calculados con el fin de poder extraer conclusiones sobre la población. Estos se suelen calcular como el valor de la estimación puntual del parámetro θ más (menos para

el extremo inferior) la estimación del error estándar del estimador multiplicado por una constante. Se asume que el estimador $\hat{\theta}$ tiene una distribución aproximadamente normal, por lo que el intervalo de confianza para un parámetro θ con un nivel de confianza $100(1 - \alpha) \%$ queda definido como:

$$\hat{\theta} \pm z_{1-\alpha/2} \times \widehat{SE}(\hat{\theta}) \quad (3.3)$$

Particularmente se trabajará con una confianza del 95 %, por lo que se sustituye en la ecuación anterior el valor de la constante $z_{1-\alpha/2}$ por 1.96.

Un intervalo con nivel de confianza del 95 % significa que, en promedio, de cada 100 muestras obtenidas bajo el mismo diseño, el intervalo contiene al verdadero valor del parámetro en 95 de ellas.

Para la construcción de los intervalos de confianza es necesario el cálculo de los errores estándar, por lo que a continuación se presentaran los métodos utilizados para el cálculo de los mismos.

3.3. Estimación de los errores estándar (SE)

Como la muestra es aleatoria, los estimadores son variables aleatorias y están, por lo tanto, sujetos a variación. Esta variación puede ser medida a partir del error estándar (SE) de la estimación. El SE es una medida de la precisión de la estimación, que se puede utilizar para realizar inferencia de la población de interés, a partir de la construcción de intervalos de confianza. Los errores estándar son calculados con el objetivo de añadir a las estimaciones puntuales una medida de calidad de las mismas. Los estimadores presentan diversas fuentes de variabilidad, dadas tanto por el diseño muestral utilizado como por los ajustes realizados a los ponderadores. Por esta razón, se busca que la estimación de los SE capte de la mejor manera posible estas variaciones.

Existen también los errores no muestrales, que no se deben a la utilización de una muestra, sino a otras causas, como por ejemplo errores en la recolección o en el

procesamiento de los datos, entre otros. Estos no serán abordados en este trabajo en particular.

El no contar con parte de la información sobre el diseño de muestreo en las bases de datos públicas de las encuestas puede llevar al usuario a obtener resultados incorrectos a la hora de estimar los SE. En el caso de la ECH, para los años anteriores a 2018 no se cuenta con la disponibilidad de información clave sobre el diseño muestral, como las unidades primarias de muestreo (UPM) y los estratos a los que pertenece cada individuo. En estos casos solo se cuenta con información sobre el hogar al que pertenece cada individuo.

No existen métodos exactos para calcular los SE en el caso de un diseño complejo como el de la ECH, ya que el estimador planteado por (Horvitz and Thompson, 1952) necesita de las probabilidades de inclusión de segundo orden, con las cuales no se cuenta. Por lo tanto, se debe optar por un método que aproxime la varianza de las estimaciones de la mejor forma posible. En este trabajo en particular se recurrió a la aproximación de una etapa planteado en (Lumley, 2010) en conjunto con la linearización de Taylor, el cual se encuentra implementado en el paquete *survey* (Lumley, 2020) del *R* (R Core Team, 2020). Este método permite calcular los SE para parámetros lineales y no lineales (Lumley, 2004).

3.3.1. Aproximación de una etapa

Este método, planteado inicialmente por (Hansen, M. H., Hurwitz, W. N., Madow, W. G., 1953) con el nombre de "*The ultimate cluster variance estimate*", es un método utilizado para la aproximación de varianzas en muestras obtenidas a partir de diseños muestrales complejos, particularmente cuando estos incluyen varias etapas de selección. El mismo consiste principalmente en utilizar únicamente la varianza de los estimadores en la primera etapa, es decir, en la selección de las UPM, sin tener en cuenta las etapas posteriores, asumiendo así que la variabilidad se encuentra mayormente en esta primera etapa de selección.

En un diseño como el de la ECH, que consta de dos etapas de selección, la va-

rianza de los estimadores tiene dos componentes: la variabilidad entre las UPM y la variabilidad dentro de las UPM. En las situaciones donde N es grande, la contribución de este segundo componente a la varianza es despreciable en comparación con el primer componente (Lohr, 2009). Por lo tanto, a la hora de estimar la varianza del estimador del total de una variable y bajo un diseño muestral como el de la ECH, se utiliza la siguiente fórmula:

$$\widehat{SE}^2(\hat{t}) = \hat{V}(\hat{t}) = \sum_{h=1}^H \frac{1}{m_h(m_h-1)} \times \sum_{j \in s_h} (\hat{t}_{jh} \times m_h - \hat{t}_h)^2 \quad (3.4)$$

Siendo $\hat{t}_{jh} = \sum_{i \in s_{jh}} w_i \times y_i$ la estimación del total de la variable y en la UPM j perteneciente al estrato h , $\hat{t}_h = \sum_{j \in s_h} \hat{t}_{jh}$ la estimación del total de la variable y en el estrato h . Mientras que m_h , como se definió anteriormente, es la cantidad de UPM pertenecientes al estrato h seleccionadas en la muestra.

En el caso de la estimación en dominios se debe remplazar la variable y , presente en la ecuación anterior 3.4, por una variable y_d que toma el valor y cuando la unidad pertenece al dominio d en cuestión, y cero cuando no pertenece al dominio. Como el tamaño de muestra en el dominio suele ser aleatorio, mediante este procedimiento se logra incluir una variabilidad extra que lo refleje.

Una de las principales ventajas que presenta este método es que utiliza directamente los expansores o pesos muestrales que se encuentran en las bases públicas de la ECH.

3.3.2. Linearización de Taylor

La mayoría de los indicadores trabajados no son lineales, ya que como se mencionó anteriormente son cocientes entre totales. La aproximación por linearización de Taylor, planteada entre otros por (Särndal et al., 2003), se basa en expresar el estimador no lineal como función de estimadores lineales. En primer lugar se utiliza el desarrollo de Taylor de primer orden para aproximar el estimador de ratio $R = t_y/t_z$:

$$\hat{R} \approx R + \frac{1}{t_z} \times \sum_{i \in s} w_i \times (y_i - Rz_i)$$

Posteriormente se aproxima la varianza de la estimación mediante la varianza de primer orden utilizando el método planteado anteriormente, sustituyendo la variable y por la variable $r_i = y_i - \hat{R}z_i$ añadiendo además el término $1/\hat{t}_z^2$.

$$\widehat{\text{SE}}^2(\hat{R}) = \hat{V}(\hat{R}) = \frac{1}{\hat{t}_z^2} \times \sum_{h=1}^H \frac{1}{m_h(m_h - 1)} \times \sum_{j \in s_h} (\hat{t}_{r,jh} \times m_h - \hat{t}_{r,h})^2 \quad (3.5)$$

Siendo $\hat{t}_z = \sum_{i \in s} w_i \times z_i$ la estimación del total de la variable z ; $\hat{t}_{r,jh} = \sum_{i \in s_{jh}} w_i \times r_i$ la estimación del total de la variable r en la UPM j perteneciente al estrato h ; y $\hat{t}_{r,h} = \sum_{j \in s_h} \hat{t}_{r,jh}$ la estimación del total de la variable r en el estrato h .

Para la estimación en un dominio en particular, se procede de la misma manera que se planteó anteriormente, sustituyendo las variables en cuestión por sus correspondientes en el dominio.

Capítulo 4

Métodos computacionales

En este capítulo se presentan en primer lugar los programas y funciones utilizados para la estimación de los indicadores, utilizando un ejemplo en concreto. Luego se realiza una introducción a las aplicaciones web Shiny y se explica el funcionamiento de la app creada en este trabajo.

4.1. Estimación de los indicadores

Para replicar el diseño muestral de la ECH y poder llevar a cabo el cálculo de los indicadores, se utiliza el software libre *R* (R Core Team, 2020), más precisamente el paquete *srvyr* (Freedman Ellis, 2020). Este paquete utiliza las funciones del paquete *survey* antes mencionado, permitiendo el uso de un estilo de sintaxis inspirado en el paquete *dplyr* (Wickham et al., 2020).

El paquete *srvyr* cuenta con funciones tales como *survey_mean* para calcular medias o proporciones, *survey_ratio* para ratios o *survey_total* para totales. Estas funciones utilizan el valor que toma el indicador en cada uno de los individuos (hogares o personas según corresponda) de la muestra y los ponderadores, dando como resultado tanto las estimaciones puntuales, como las estimaciones de los errores estándar, utilizando las ecuaciones y metodologías descritas en este documento para realizar los cálculos.

CAPÍTULO 4. MÉTODOS COMPUTACIONALES

A continuación se presenta un breve ejemplo de cómo se utilizan las funciones de esta librería, para el caso concreto del indicador *Utilización de computadora el último mes por sexo*.

```
pr<-echP%>% as_survey_design(ids=upm_fic, weight=pesoano
                             , strata=estrato)

diseno<- pr
filtro<- "e27>5"
variable<- "e61==1"

ind<-diseno%>%
  filter(eval(parse(text=filtro)))%>%
  group_by(dpto, e26)%>%
  summarise(estimacion.tot= survey_mean(eval(parse(text=variable))))
```

En primer lugar se crea el objeto de diseño mediante la función *as_survey_design*, donde los argumentos que se le indican son las unidades primarias de muestreo (*upm_fic*), los ponderadores (*pesoano*) y los estratos (*estrato*). Todas estas son variables presentes en las base de datos de la ECH 2019. Luego se crean los objetos *filtro* y *variable*, el filtro define la población objetivo, que en este caso son las personas mayores de 5 años, mientras que la variable utilizada es *e61 == 1*, la cual es la variable de la ECH, que hace referencia a la utilización de computadora en el último mes.

Las líneas de código siguientes son la base de las funciones creadas para el cálculo de los indicadores, donde en este caso se le indica el objeto de diseño, se le aplica el filtro necesario, se agrupa por departamento y por sexo (*e26*) y se estima la media de la variable indicada para cada sexo a nivel de cada departamento, dando como resultado un objeto (*data frame*) con las estimaciones puntuales y errores estándar para el indicador deseado.

Tanto para el filtro como para la variable se utilizan las funciones *eval* y *parse*, la función *parse* convierte texto en expresiones, mientras que la función *eval* evalúa expresiones.

Para la simplificación del cálculo de los indicadores se crearon 9 funciones que varían dependiendo de los individuos involucrados (personas y hogares), la forma en la que se calcula cada indicador y las desagregaciones necesarias. Todas estas funciones estiman el indicador tanto a nivel de cada departamento, como de todo el país y tienen como salida una tabla (data frame) con los resultados.

Todas las funciones necesitan de algunos insumos para estimar los indicadores correspondientes, estos son:

- El diseño muestral.
- El filtro a aplicar (de ser necesario), que determina la población objetivo para la cual el indicador será calculado.
- Variable(s) de interés: variable(s) necesaria(s) para el cálculo del indicador.
- Vector con categorías de la variable de interés (en caso de ser necesario).

Las primeras siete funciones se utilizan para los indicadores que se obtienen a partir de la base de personas de la ECH, mientras que las últimas dos trabajan con la base de hogares.

La **funcion1** se utiliza para los indicadores que necesitan del cálculo de la proporción de individuos que cumplen determinada condición (por ejemplo quienes tienen estudios terciarios), desagregando además por sexo.

La **funcion2** es utilizada para los indicadores en los cuales es necesario calcular la proporción de distintas categorías de una variable, y también se desagrega por sexo.

Las funciones **funcion3**, **funcion4** y **funcion5** son similares, calculan proporciones al igual que en la función 1, pero desagregando en tramos de edad en lugar

de por sexo. La diferencia entre estas tres funciones es que desagregan en distintos tramos etarios.

La `funcion6` se utiliza únicamente para la estimación del indicador *Promedio de años de educación de las personas de 25 años y más por sexo*. La variable en este caso es cuantitativa, se promedia y se desagrega por sexo.

La `funcion7` se utiliza para los indicadores que hacen referencia a tasas brutas, es decir, que necesitan del cálculo de un ratio.

Los indicadores *Población total por condición de actividad por sexo* y *Personas por tipo de atención en salud*, fueron calculados de manera particular, sin el uso de ninguna función.

Las funciones `funcion8` y `funcion9`, como se mencionó anteriormente, son las utilizadas para la estimación de los indicadores que se obtienen a partir de la base de hogares de la ECH. La `funcion8` se utiliza para los indicadores que requieren el cálculo de la proporción de hogares que cumplen con determinada condición, mientras que en la `funcion9`, la variable en cuestión presenta más de 2 categorías.

El indicador *Hogares por tipo de relación con la vivienda* fue estimado de manera particular, sin el uso de ninguna función.

Se crean además funciones para el cálculo de los intervalos de confianza de las estimaciones, funcionando de la misma manera que las funciones mencionadas anteriormente, pero obteniendo una salida diferente, nueve de ellas dan como salida una tabla con los extremos inferiores de los intervalos, mientras que las otras nueve hacen lo propio con los límites superiores.

El cálculo del indicador presentado anteriormente, una vez creadas las funciones, se reduce a la siguiente línea de código.

```
funcion1(pr, "e27>5", "e61==1")
```

Este indicador es estimado en base a personas y con desagregación por departamento y por sexo. Se utiliza la función `funcion1`, el diseño utilizado `pr`, el filtro `e27 > 5` y la variable `e61 == 1` son los presentados anteriormente.

4.1.1. Descripción de los diseños

El argumento *diseño muestral* que utilizan las funciones creadas para el cálculo de los indicadores, depende de la información disponible con la que cuente el usuario. En las bases públicas del INE no se cuenta con información del estrato ni de la UPM a la que pertenece el individuo, por lo que en la app aparecen dos opciones. La opción *pública* se utiliza cuando no se cuenta con esta información, mientras que en caso de contar con la misma se utiliza la opción *estratos+UPM*.

En caso de no contar con la información de estratos y UPM, los cálculos de los intervalos de confianza son incorrectos, ya que esta información es esencial para poder realizar una aproximación correcta de los SE y así poder construir los intervalos de confianza. La única información con la que se cuenta para realizar los cálculos en este caso es el hogar al que pertenecen los individuos, es decir se tiene en cuenta cuando dos individuos pertenecen al mismo hogar, pero no así cuando dos individuos pertenecen al mismo estrato o a la misma UPM. Los hogares que pertenecen a un mismo conglomerado (UPM) suelen comportarse de manera similar en muchos de los tópicos abordados en la ECH, por lo que no tener en cuenta esta información hace se tomen a los 42000 hogares como unidades primarias y por lo tanto la muestra parezca más grande de lo que realmente es.

Cabe aclarar que el INE hizo pública esta información para los años 2018 y 2019.

4.2. Aplicación web utilizando Shiny

Shiny (Chang et al., 2020) es una librería que permite crear fácilmente aplicaciones web interactivas a partir de R. La interactividad de estas aplicaciones permite la manipulación de los datos sin la necesidad de manipular el código, es decir que, en caso de estar disponible en la web, puede ser fácilmente utilizada por personas que no tengan conocimientos de R.

Una Shiny app contiene al menos dos componentes: un script para la interfaz del usuario (UI) donde se controla el diseño de la aplicación; y un script para realizar

los cálculos necesarios (server). En el server se debe especificar cómo convertir los valores de entrada (inputs) en resultados (outputs).

Estas aplicaciones cuentan con programación reactiva. Esto significa que al cambiar los valores de entrada (inputs) se volverán a ejecutar las partes del código de R correspondientes, lo que a su vez hará que se actualicen las salidas (outputs) modificadas, es decir, que la modificación de un valor reactivo llevará automáticamente a todas las expresiones reactivas que dependen directa o indirectamente de este valor a ejecutarse nuevamente.

Se cuenta además con widgets pre-construidos que facilitan la construcción de aplicaciones bonitas e interactivas. Un widget es un elemento web con el cual los usuarios pueden interactuar, enviando mensajes a la aplicación Shiny.

4.2.1. Monitor de indicadores socioeconómicos

Con el objetivo de poder obtener resultados sobre distintos indicadores obtenidos a partir de la ECH, pudiendo visualizarlos interactivamente a partir de tablas y gráficos, y con la posibilidad de contar con una medida de precisión de las estimaciones, se crea esta aplicación web a partir de Shiny.

La aplicación procesa los datos de la ECH. Particularmente se realizó en base a la encuesta del año 2019, pero también funciona perfectamente para años anteriores y está pensada para que funcione también para los próximos años, en caso de que la encuesta no tenga cambios abruptos en su formato y estructura. Para el cálculo de los indicadores se utilizan un grupo de variables de la ECH, lo que hace a la app dependiente de esas variables, es decir, que en caso de que esas variables cambien de nombre o dejen de estar disponibles, la app deberá ser actualizada. Tanto en la aplicación como en los anexos A.2 de este informe, estará disponible un listado con las variables utilizadas.

En esta aplicación en particular, previo a la creación del UI y el server, se crea un archivo con nombre *global.R*. Allí se cargan las librerías a utilizar, los datos espaciales necesarios para los gráficos de mapas, el listado de variables utilizadas de

la ECH y se crean las funciones que se utilizan para el cálculo de los indicadores. Esta parte del código se corre solo una vez al iniciar la app.

4.2.2. Interfaz del usuario

Para una mejor visualización se utilizó la librería *shinydashboard*, que cuenta con un conjunto de funciones diseñadas para crear *HTML* para generar un tablero (*dashboard*). La UI cuenta con tres partes, el encabezado (*dashboardHeader*), la barra lateral (*dashboardSidebar*) y el contenido (*dashboardBody*).

Dentro del encabezado solo aparece el título y un link que lleva directamente a la página del INE donde están disponibles las bases de datos de la ECH.

La barra lateral 4.1 es utilizada para cargar las bases de datos y para seleccionar la sección que se desee visualizar, ya sea la pestaña de introducción o las correspondientes a las diferentes categorías de los indicadores. Para cargar los datos se utiliza la función *fileInput* y se deben cargar por separadas las bases de personas y de hogares de la ECH y el archivo que contiene la información acerca de los estratos y UPM (en caso de contar con este). También se incluye un widget mediante la función *selectInput* para que el usuario pueda especificar si las bases utilizadas son las públicas o si se cuenta con la información de estratos y UPM.

En el body se especifica el contenido de cada una de las secciones. En primer lugar se encuentra la ventana de inicio 4.2, que fue creada para realizar una introducción a la app, y cuenta con dos pestañas creadas con la función *tabPanel*. En la primera de estas se encuentra un manual de uso de la aplicación, donde se explica su utilidad y los pasos a seguir para el correcto funcionamiento de la misma. Por otra parte, en la segunda pestaña se encuentra un listado con las variables de las bases de datos de la ECH que son utilizadas en la app, las cuales, como se mencionó anteriormente, son fundamentales para el funcionamiento de la aplicación.

La sección de los indicadores cuenta con 3 pestañas. En la primera de estas 4.3 se selecciona el indicador que se quiere calcular y se pueden visualizar dos tipos de gráficos para los resultados. Un gráfico de mapa donde, en caso de que correspon-

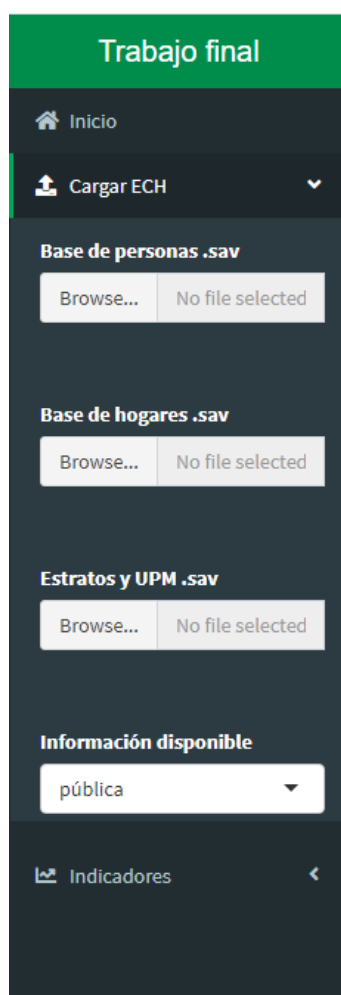


Figura 4.1: Barra lateral.

da, se puede seleccionar la categoría a graficar, y un diagrama de barras (apiladas, agrupadas o simple según el caso). Para poder observar ambos gráficos se debe hacer click en el botón **Go** creado mediante *actionButton*. El gráfico de mapa es una herramienta útil para poder visualizar los resultados, en este caso, a nivel departamental, pudiendo observar posibles aspectos geográficos. Mientras el gráfico de barras se utiliza para resumir un conjunto de datos por categorías, utilizando las barras simples cuando la única categorización es departamental, agrupadas cuando se presenta además otra variable con diferentes categorías (por ejemplo el sexo), y apiladas cuando además de presentar otra variable categórica, estas categorías suman el 100 % dentro de cada grupo (departamento).

4.2. Aplicación web utilizando Shiny

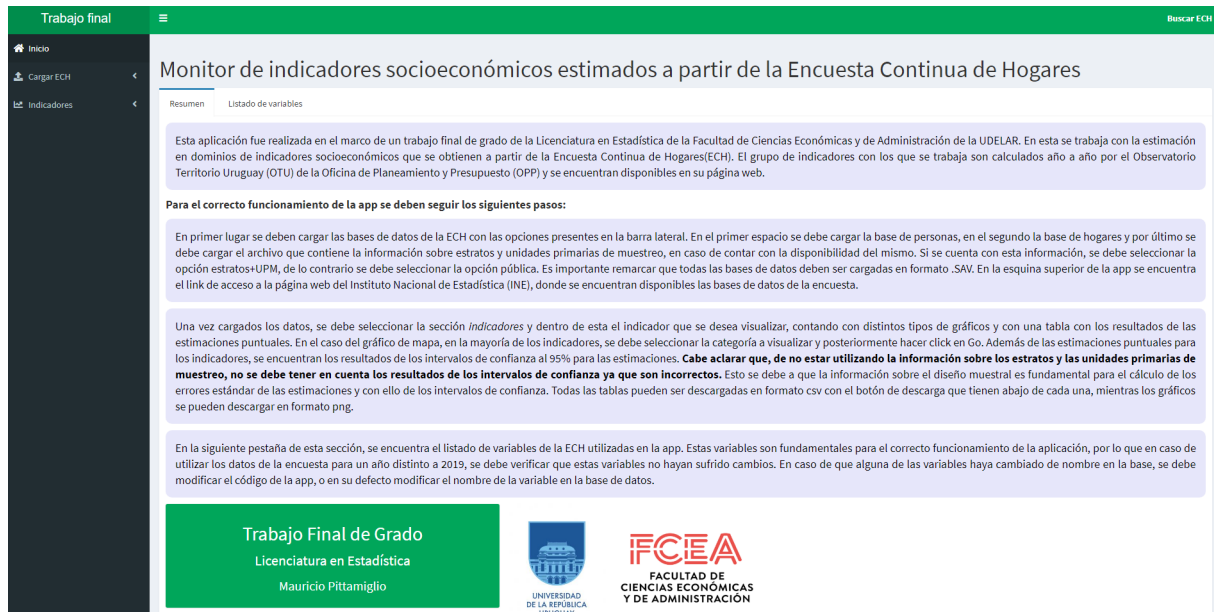


Figura 4.2: Ventana inicial de la app.

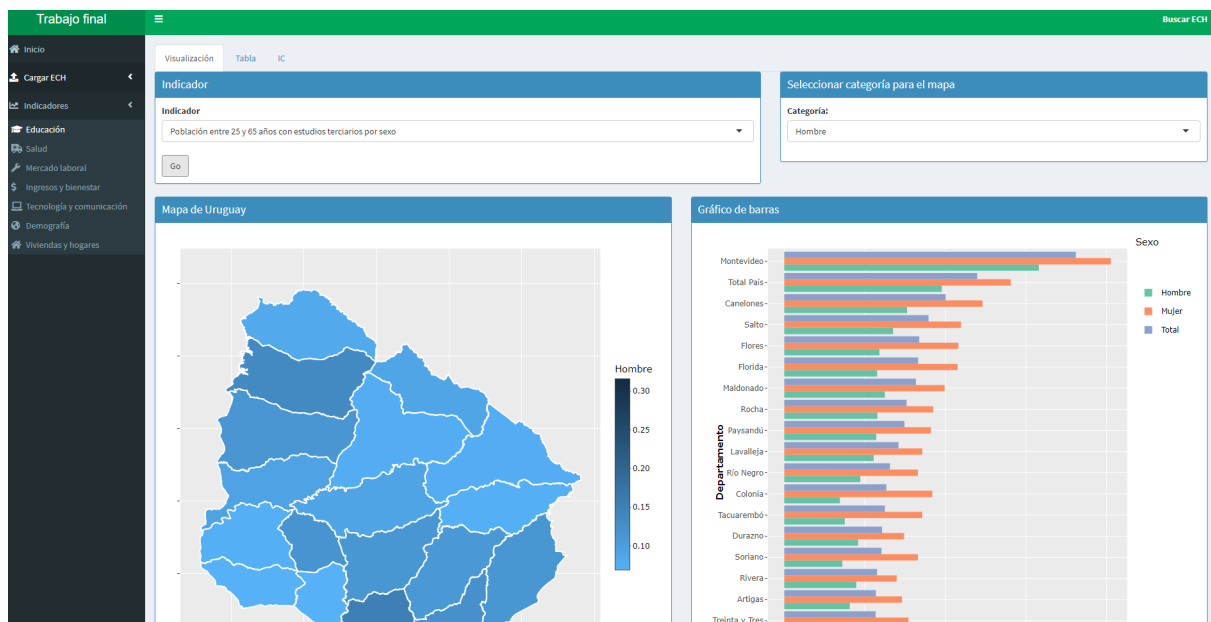


Figura 4.3: Sección indicadores.

En la segunda pestaña se puede observar una tabla para las estimaciones puntuales del indicador seleccionado y un botón de descarga (*downloadButton*) para poder descargar dicha tabla en formato csv. A partir de una tabla se pueden visualizar los resultados en formato de filas y columnas, permitiendo al usuario filtrar valores.

Mientras la tercera pestaña (IC) cuenta con un gráfico y una tabla donde se visualiza la estimación puntual del indicador en cuestión, con su respectivo intervalo de confianza al 95 %, contando además con un seleccionador de la categoría a observar en caso de ser necesario. Esta última es posiblemente la forma más completa de visualizar los resultados, ya que se presenta en el mismo gráfico (o tabla) el valor de la estimación puntual del indicador junto con una medida de calidad de la estimación como lo es el intervalo de confianza.

4.2.3. Comportamiento reactivo

La función *server* toma dos parámetros: *input* y *output*. El argumento *input* es un objeto similar a una lista que contiene todos los elementos de entrada enviados por el navegador. Mientras que el argumento *output* es similar al anterior con la principal diferencia de que es usado para enviar objetos de salida y no para recibir objetos de entrada.

En esta parte del código se controla el comportamiento de la app. En primer lugar se crean todos los elementos reactivos (se guardan las bases de datos cargadas, los diseños y todos los indicadores) utilizando la función *reactive*. Las expresiones reactivas tienen dos propiedades importantes: no funcionan hasta que no se las llama y quedan almacenadas en el caché, es decir, no hace ningún trabajo ni la segunda ni ninguna de las siguientes veces que se las llama, porque queda almacenado en el caché el resultado anterior. Luego se crean todos los componentes de salida con funciones *render*, las tablas con *renderDT* y los gráficos con *renderPlotly*. Las funciones *render* cumplen con dos objetivos: establecer un contexto reactivo especial que rastrea automáticamente qué entradas usa la salida y convierte estas salidas del código de *R* en *HTML* adecuado para mostrar en una página web (Wickham, 2020).

Plotly (Sievert, 2020) hace los gráficos interactivos, cuenta con varias opciones para poder visualizarlos de manera más precisa y permite descargarlos como imagen.

Para el caso de la selección de la categoría a visualizar en el gráfico de mapa y el gráfico con los IC, se utilizan las funciones *uiOutput* y *renderUI*. La primera de estas inserta un marcador de posición en la interfaz de usuario, es decir que el código se ejecuta cuando se inicia la aplicación pero deja un espacio para que el código del servidor (server) pueda llenar luego. Mientras que *renderUI* llena ese espacio con UI generada en el servidor. Esta técnica permite regenerar la interfaz del usuario mientras la app está corriendo, y se utiliza en este caso ya que las categorías de los indicadores no son siempre las mismas, sino que varían según el indicador seleccionado.

Capítulo 5

Ejemplo de aplicación

En esta sección se presenta el procedimiento aplicado y los resultados obtenidos para el caso de dos indicadores en particular, la **tasa de desempleo por sexo** y **Hogares por presencia de lugar para cocinar**. Utilizando la ECH del 2019 y la información sobre los estratos y unidades primarias de muestreo (UPM).

5.1. Tasa de desempleo

La tasa de desempleo es uno de los indicadores más importantes que brinda la ECH. El INE calcula este indicador todos los años a nivel departamental e incluso desagregando también por sexo.

Este indicador hace referencia a la proporción de personas desocupadas por sexo, donde, desocupada es toda persona que durante el período de referencia de la encuesta, no está trabajando por no tener empleo, que lo busca activamente, y están disponibles para empezar a trabajar. Se calcula como $\frac{\text{Numero de personas desocupadas}}{\text{Poblacion Economicamente Activa}}$.

Pasando al cálculo a partir de las funciones, el código es el siguiente:

```
funcion1(pr, "e27>13 & activo==1", "desocupados==1")
```

Al ser un indicador calculado en base a personas, necesitando del cálculo de una proporción y desagregando por sexo, se utiliza la *funcion1*. El diseño utilizado es *pr*,

CAPÍTULO 5. EJEMPLO DE APLICACIÓN

el filtro $e27 > 13$ y $activo == 1$ indica que se calcula para los mayores de 13 años y la variable $activo == 1$, es una variable creada a partir de otras variables de la ECH, que toma valor 1 cuando el individuo pertenece a la población económicamente activa. Finalmente la variable utilizada es $desocupados == 1$ la cual también es una variable creada a partir de otras variables de la ECH y toma valor 1 cuando el individuo está desocupado y 0 si no lo está.

Para la obtención de los resultados de este indicador en la app, en primer lugar se cargan las bases de datos necesarias, disponibles en la página web del INE. Primero se debe cargar la base de personas, luego la de hogares y en tercer lugar, el archivo con la información sobre estratos y UPM.

Una vez cargadas las bases se selecciona dentro de la sección de indicadores, la categoría *Mercado laboral* donde se encuentra este indicador. Una vez dentro de esta sección, se selecciona el indicador que se desea obtener.

En la misma pestaña se observan como resultados los dos tipos de gráficos antes mencionados, los cuales se presentan en las figuras 5.1 y 5.2.

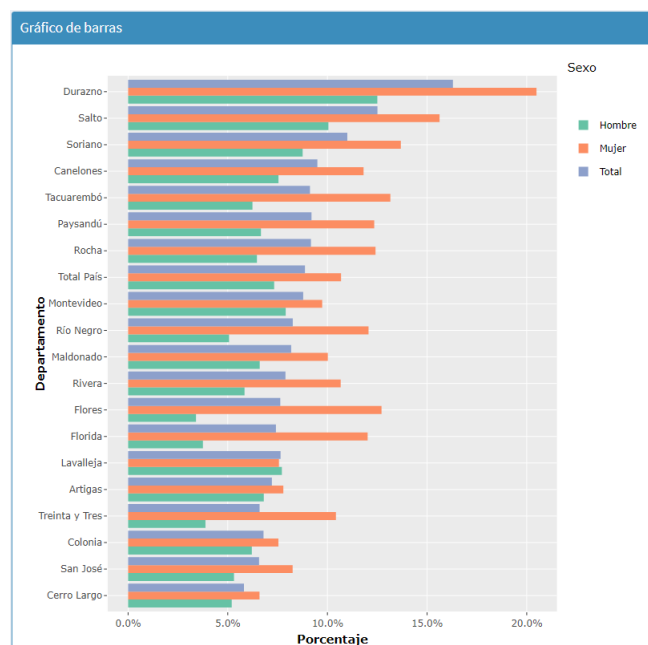


Figura 5.1: Gráfico de barras de la tasa de desempleo para el año 2019 a nivel departamental y de todo el país desagregando por sexo.

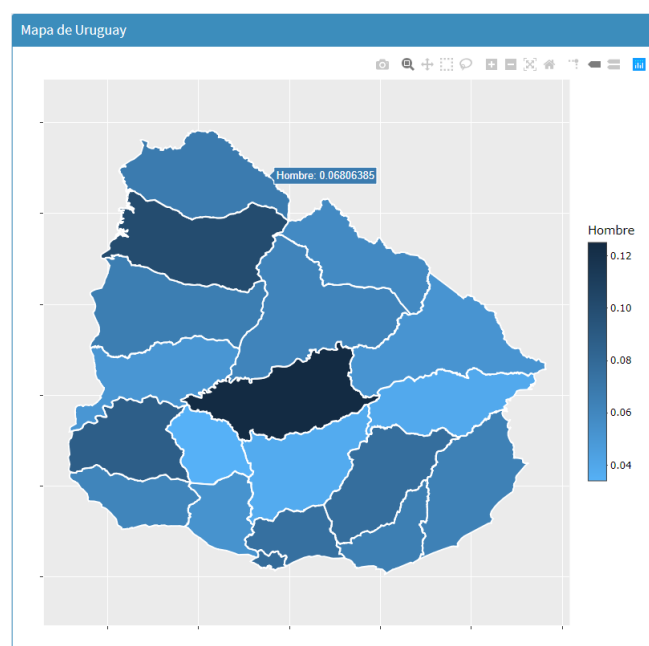


Figura 5.2: Tasa de desempleo a nivel departamental para la categoría Hombres.

En la Figura 5.1 los departamentos están ordenados por el valor que toman, más precisamente por el promedio de los tres valores (Hombre, Mujer y Total) correspondientes al departamento, por lo que podemos observar que los departamentos con mayor tasa de desempleo son Durazno y Salto, mientras Cerro Largo y San José son los que presentan menor desempleo. También podemos observar que el desempleo afecta más a las mujeres en prácticamente todos los departamentos.

Para el gráfico de mapa se debe seleccionar la categoría a graficar, en este caso las opciones son *Hombre*, *Mujer* o *Total*, de las cuales se selecciona la categoría *Hombre*.

En este gráfico presentado en la Figura 5.2 se pueden visualizar y analizar además aspectos geográficos, como comparar la tasa de desempleo en departamentos cercanos. Por ejemplo podemos observar que los departamentos ubicados en el sureste del país se comportan de manera similar en este caso.

Como se puede observar en la Figura 5.2, al colocar el cursor sobre uno de los departamentos, el gráfico nos brinda el valor de la estimación para dicho departamento. Esto también es posible en el diagrama de barras presentado y es una de las

CAPÍTULO 5. EJEMPLO DE APLICACIÓN

ventajas del uso de gráficos interactivos creados con la librería *plotly*.

En la segunda pestaña de esta sección se puede observar una tabla como la presentada en la Figura 5.3, que contiene las estimaciones puntuales del indicador seleccionado, a nivel departamental y nacional.

Search: <input type="text"/>			
depto	Hombre	Mujer	Total
Montevideo	0.0791	0.0974	0.0878
Artigas	0.0681	0.0779	0.0721
Canelones	0.0754	0.1181	0.0950
Cerro Largo	0.0520	0.0659	0.0581
Colonia	0.0621	0.0754	0.0679
Durazno	0.1250	0.2049	0.1630
Flores	0.0340	0.1271	0.0763
Florida	0.0375	0.1202	0.0742
Lavalleja	0.0771	0.0757	0.0765
Maldonado	0.0660	0.1002	0.0818
Paysandú	0.0666	0.1235	0.0920
Rio Negro	0.0506	0.1206	0.0826
Rivera	0.0584	0.1066	0.0789
Rocha	0.0646	0.1241	0.0917
Salto	0.1005	0.1562	0.1251
San José	0.0531	0.0825	0.0657
Soriano	0.0876	0.1368	0.1100
Tacuarembó	0.0624	0.1316	0.0912
Treinta y Tres	0.0388	0.1043	0.0659
Total País	0.0733	0.1068	0.0887
Descargar			

Figura 5.3: Tabla con las estimaciones puntuales de la tasa de desempleo para los 19 departamentos y para todo el país en el año 2019.

Por último, en la pestaña *IC*, se observan un gráfico y una tabla donde se presenta la estimación puntual del indicador junto con su respectivo intervalo de confianza. Se presenta el gráfico y la tabla correspondientes para la tasa de desempleo en hombres en las Figuras 5.4 y 5.5 respectivamente.

En la Figura 5.4 el orden de los departamentos está determinado por el valor de la estimación puntual del indicador en el mismo. Podemos observar que la mayoría de los intervalos se solapan, por lo que las diferencias entre estos departamentos no son estadísticamente significativas. También se observa que los intervalos son menos amplios para el total país y los departamentos más poblados (Montevideo y Canelones), lo cual es esperable debido al mayor tamaño de muestra en estos casos.

5.1. Tasa de desempleo

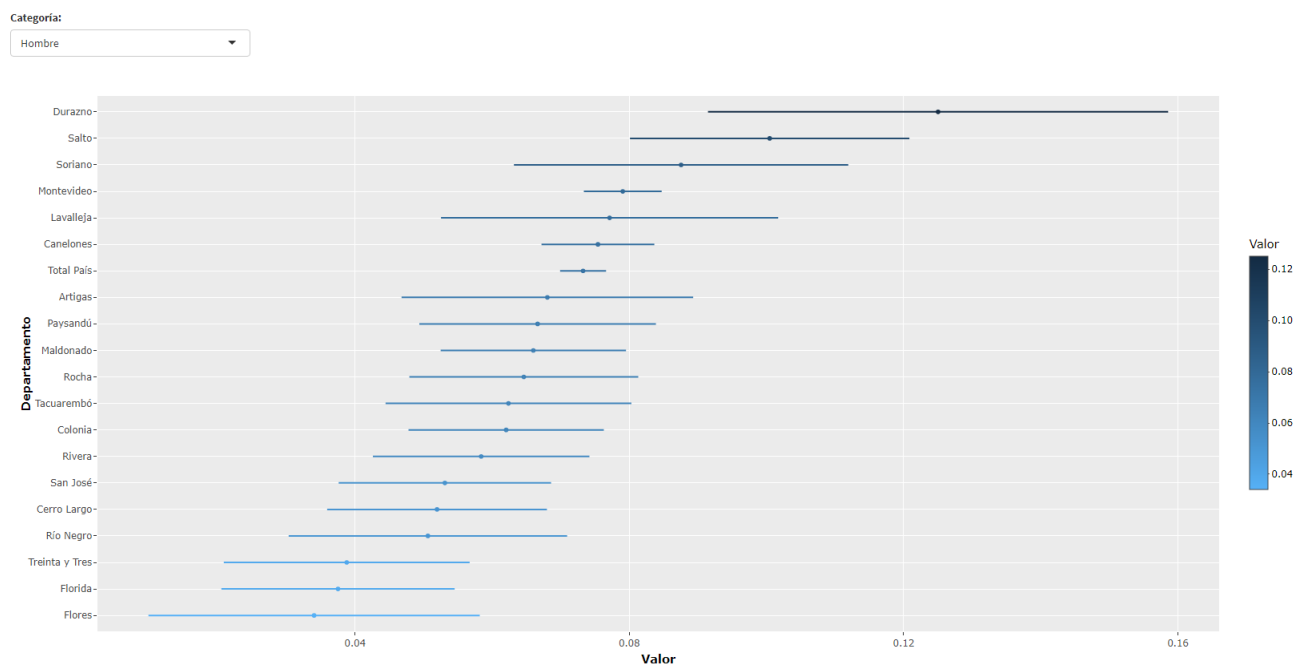


Figura 5.4: Estimación de la tasa de desempleo en hombres para los 19 departamentos y para todo el país en el año 2019, con los respectivos intervalos de confianza al 95 %.

dpto	Valor	Inferior	superior
Montevideo	0.0791	0.0734	0.0847
Artigas	0.0681	0.0468	0.0893
Canelones	0.0754	0.0672	0.0837
Cerro Largo	0.0520	0.0360	0.0680
Colonia	0.0621	0.0478	0.0763
Durazno	0.1250	0.0915	0.1586
Flores	0.0340	0.0099	0.0582
Florida	0.0375	0.0205	0.0545
Lavalleja	0.0771	0.0526	0.1017
Maldonado	0.0660	0.0525	0.0795
Paysandú	0.0666	0.0494	0.0839
Río Negro	0.0506	0.0303	0.0709
Rivera	0.0584	0.0426	0.0742
Rocha	0.0646	0.0479	0.0813
Salto	0.1005	0.0801	0.1209
San José	0.0531	0.0376	0.0686
Soriano	0.0876	0.0632	0.1119
Tacuarembó	0.0624	0.0445	0.0803
Treinta y Tres	0.0388	0.0209	0.0567
Total País	0.0733	0.0699	0.0766

Figura 5.5: Tabla con la estimación de la tasa de desempleo en hombres para los 19 departamentos y para todo el país en el año 2019, con los respectivos intervalos de confianza al 95 %.

5.2. Hogares por presencia de lugar para cocinar

Por otra parte, el indicador **Hogares por presencia de lugar para cocinar** hace referencia a la proporción de hogares que cuentan con lugar apropiado para cocinar en privado, con lugar apropiado compartido con otros hogares, o que no cuentan con lugar apropiado y se calcula como $\frac{\text{Numero de hogares segun presencia de lugar para cocinar}}{\text{Total hogares}}$.

```
funcion9(ph, "d19", c("dpto", "Lugar privado", "Compartido",  
                    "No tiene"), c(1,2,3))
```

Este indicador se calcula sobre la base de hogares, se utiliza la *funcion9*, el diseño *ph*. No tiene filtro ya que se calcula sobre todos los hogares. La variable utilizada es *d19* la cual es una variable presente en la ECH que hace referencia al lugar para cocinar y toma valor 1 cuando tiene lugar privado, 2 cuando es compartido con otros hogares y 3 cuando no tiene. Esta información es la que se le brinda a la función en los otros dos insumos.

En la app este indicador se encuentra en la sección *Viviendas y hogares*, por lo que de igual manera que para el indicador anterior, se ingresa a la sección, se selecciona el indicador deseado y se observan los resultados.

La diferencia en este indicador es que se presenta un gráfico de barras apiladas (5.6), en lugar de agrupadas como era el caso de la tasa de desempleo, ya que en este caso se desea observar la proporción que representa cada categoría de la variable en cuestión dentro de cada departamento. Las categorías acumulan el 100 %, por lo que de esta forma se pueden observar las diferencias relativas entre las cantidades en cada grupo.

En la Figura 5.6 se observa que en todos los departamentos predomina la presencia de lugar privado para cocinar en los hogares. En Río Negro es donde la categoría *Compartido* tiene mayor frecuencia, mientras los departamentos de Durazno, Paysandú y Salto son aquellos en los que los hogares que no tienen lugar para cocinar

5.2. Hogares por presencia de lugar para cocinar

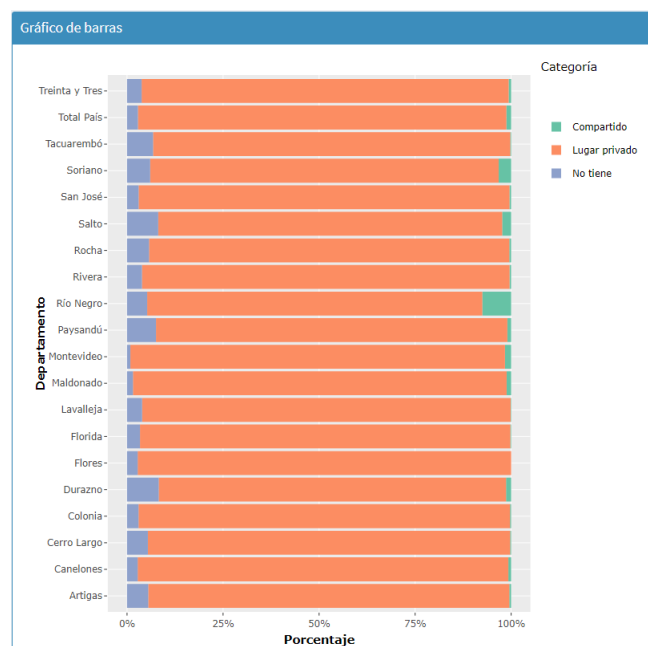


Figura 5.6: Gráfico de barras apiladas al 100 % de la proporción de hogares por presencia de lugar para cocinar por departamento para el año 2019.

representan un porcentaje mayor.

El gráfico de mapa es similar al presentado para la tasa de desempleo, donde también se puede seleccionar la categoría a visualizar. La visualización de las tablas de estimaciones puntuales e intervalos de confianza también son similares al caso del indicador anterior.

En el caso de este indicador, como se observa en la Figura 5.7, también hay intervalos que se solapan, mientras que el departamento con el intervalo más amplio es Río Negro (0.843 - 0.90).

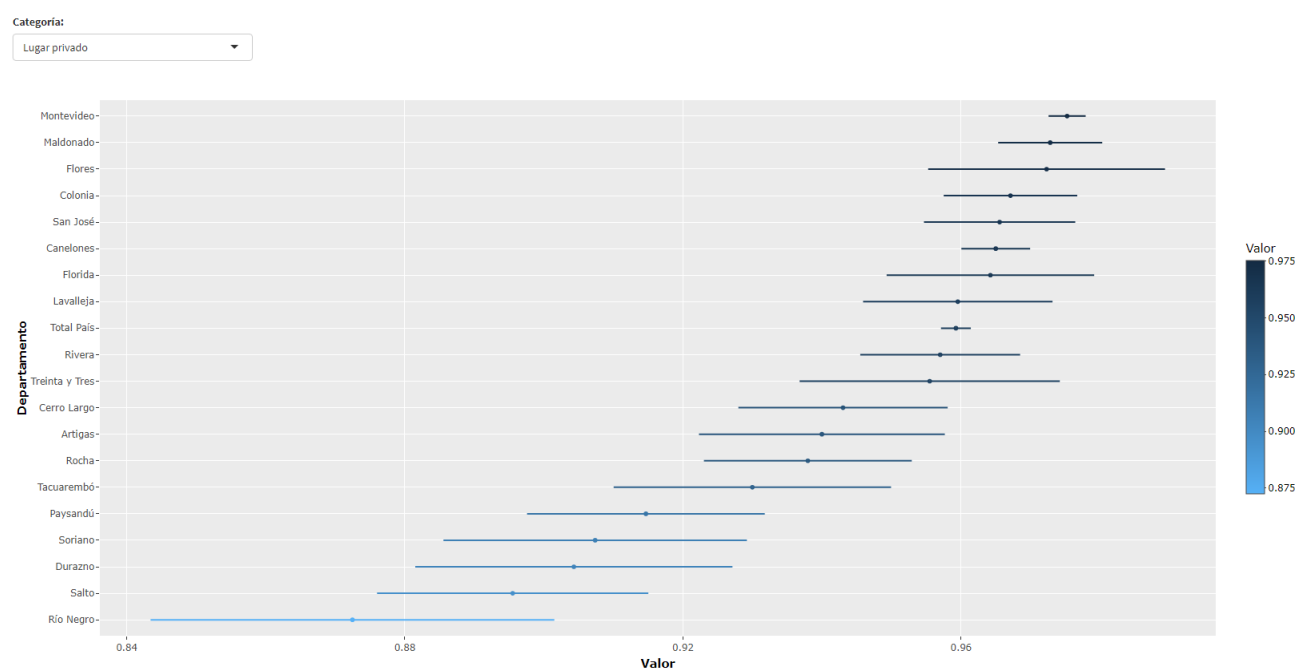


Figura 5.7: Hogares con lugar privado para cocinar para los 19 departamentos y para todo el país en el año 2019, con los respectivos intervalos de confianza al 95 %.

Capítulo 6

Discusión

La Encuesta Continua de Hogares permite obtener estimaciones sobre una gran cantidad de indicadores que abordan distintas dimensiones. Cada año el INE calcula una determinada cantidad de indicadores, en su mayoría a nivel de todo el país y desagregando en Montevideo e Interior, añadiendo además algunas medidas de calidad de las estimaciones, como lo son los intervalos de confianza o coeficientes de variación.

Cualquier usuario puede obtener sus propias estimaciones puntuales sobre distintos indicadores a partir de la ECH, ya que para ello es necesario contar únicamente con los pesos muestrales, los cuales, se encuentran incluidos en las bases de datos públicas. Sin embargo resulta más complejo el poder obtener ciertas medidas de calidad de las estimaciones.

Como se mencionó en este trabajo, las medidas de calidad se obtienen a partir del cálculo de los errores estándar, y para calcular estos errores se necesitan tener en cuenta todas las fuentes de variación que afectan a las estimaciones. Se debe tener conocimiento sobre el proceso de selección de la muestra y contar con la disponibilidad de esta información en las bases públicas, en este caso la información sobre estratos y unidades primarias de muestreo.

Otra de las principales dificultades es la utilización de programas estadísticos para poder trabajar con estos datos. Allí se encuentra una de las principales ventajas

que tiene la App de Shiny presentada en este trabajo, ya que no se necesita del conocimiento de ningún programa estadístico para su utilización. Esta App permite obtener en el momento, estimaciones puntuales sobre una batería de indicadores y sus respectivos intervalos de confianza, con solo cargar las bases de datos de la ECH del año que se desee.

Otra de las ventajas que presenta esta aplicación, es el cálculo de una gran cantidad de indicadores en diferentes dominios, ya sean aperturas departamentales, variables como el sexo o diferentes tramos etarios.

6.1. Limitaciones y trabajos a futuro

Una de las principales limitaciones que puede tener esta aplicación, es la dependencia sobre las variables de la ECH, es decir, el hecho de que si las variables utilizadas cambian de nombre o dejan de aparecer en encuestas de años próximos, la app deberá ser modificada para su correcto funcionamiento.

Otras mejoras que puede tener la app son por ejemplo: otras medidas de calidad de las estimaciones, como lo son coeficientes de variación o tamaños de muestra efectivos; obtener estimaciones en dominios más pequeños, como por ejemplo separar a las capitales departamentales del resto del departamento, siempre y cuando los tamaños de muestra sean lo suficientemente grandes para brindar estimaciones confiables; agregar más indicadores o dejar más libertad al usuario para que pueda calcular sus propios indicadores.

Independientemente de las posibles mejoras que pueda tener la app, siempre va a quedar a cargo del usuario el apropiado uso de la misma y la correcta interpretación de los resultados.

Un posible trabajo a futuro, es el de comparar los errores estándar de las estimaciones de algunos indicadores con los dos diseños planteados anteriormente, es decir, con el diseño completo con la información sobre los estratos y las UPM, y el diseño sin esta información, con el objetivo de poder observar cuánto difieren los cálculos

en ambos casos para diferentes indicadores.

Bibliografía

Chang, W., Cheng, J., Allaire, J., Xie, Y., and McPherson, J. (2020). *shiny: Web Application Framework for R*. R package version 1.4.0.2.

Coimbra, A., Goyeneche, J.J., Ferreira, J.P., Pittamiglio, M. (2020). Revisión de indicadores del observatorio territorio uruguay estimación para Áreas pequeñas. El objetivo de este trabajo es realizar una revisión crítica de los indicadores que reporta el OTU y generar una serie de instrumentos que garanticen la validez y confiabilidad estadística de los mismos.

Ferreira, J. P. (2020). Visualización de datos en encuestas públicas sobre indicadores de mercado laboral. Master's thesis, Universidad europea Miguel de Cervantes.

Freedman Ellis, G. (2020). *srvyr: 'dplyr'-Like Syntax for Summary Statistics of Survey Data*. R package version 0.3.9.

Gutiérrez, H. A. (2016). *Estrategias de muestreo: diseño de encuestas y estimación de parámetros*. Ediciones de la U.

Hansen, M. H., Hurwitz, W. N., Madow, W. G. (1953). *Sample Survey Methods and Theory*, volume I. John Wiley and Sons, Inc.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260).

INE. (2020). *Ficha técnica Encuesta Continua de Hogares*. División Estadísticas

BIBLIOGRAFÍA

- Sociodemográficas, Departamento Encuesta Continua de Hogares, Montevideo, Uruguay.
- Lohr, S. L. (2009). *Sampling: design and analysis*. Nelson Education., second edition.
- Lumley, T. (2004). Analysis of Complex Survey Samples. *Journal of Statistical Software*, 9(i08).
- Lumley, T. (2020). survey: analysis of complex survey samples. R package version 4.0.
- Lumley, T. S. (2010). *Complex Surveys: A Guide to Analysis Using R*. Wiley Publishing.
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Särndal, C.-E., Swensson, B., and Wretman, J. (2003). *Model assisted survey sampling*. Springer Science & Business Media.
- Sievert, C. (2020). *Interactive Web-Based Data Visualization with R, plotly, and shiny*. Chapman and Hall/CRC.
- Wickham, H. (2020). *Mastering Shiny: Build Interactive Apps, Reports, and Dashboards Powered by R*. O'Reilly Media, Inc, USA.
- Wickham, H., François, R., Henry, L., and Müller, K. (2020). *dplyr: A Grammar of Data Manipulation*. R package version 0.8.5.

Apéndice A

Anexos

A.1. Listado de indicadores

Se presenta el listado completo de los indicadores trabajados en la app.

Tabla A.1: Listado de indicadores

Indicador	Categoría	Función	Desagregación	Base
Tasa de analfabetismo en población de 15 años o más	Educación	1	Dpto - Sexo	Personas
Adolescentes de 12 a 17 años que asisten a establecimientos educativos	Educación	1	Dpto - Sexo	Personas
Personas mayores de 15 años que completaron media básica general o técnica	Educación	1	Dpto - Sexo	Personas
Personas de 18 años o más que completaron segundo ciclo de educación media	Educación	1	Dpto - Sexo	Personas
Población entre 25 y 65 años con estudios terciarios	Educación	1	Dpto - Sexo	Personas

APÉNDICE A. ANEXOS

Tasa neta de asistencia de 0 a 2 años en educación inicial	Educación	1	Dpto - Sexo	Personas
Tasa neta de asistencia de 3 a 5 años en educación preescolar	Educación	1	Dpto - Sexo	Personas
Tasa neta de asistencia de 6 a 11 años en educación primaria	Educación	1	Dpto - Sexo	Personas
Tasa neta de asistencia de 12 a 17 años en educación media	Educación	1	Dpto - Sexo	Personas
Personas mayores a 25 años por máximo nivel educativo alcanzado	Educación	2	Dpto - Sexo	Personas
Ocupados por máximo nivel educativo alcanzado	Educación	2	Dpto - Sexo	Personas
Tasa de analfabetismo en mayores de 15 años	Educación	3	Dpto - Tramos de edad	Personas
Promedio de años de educación de las personas de 25 años o más	Educación	6	Dpto - Sexo	Personas
Tasa bruta de asistencia de 3 a 5 años en educación preescolar	Educación	7	Dpto	Personas
Tasa bruta de asistencia de 6 a 11 años en educación primaria	Educación	7	Dpto	Personas

A.1. Listado de indicadores

Tasa bruta de asistencia de 12 a 17 años en educación media	Educación	7	Dpto	Personas
Jóvenes de 14 a 24 años que no estudian ni trabajan	Educación / Mercado Laboral	1	Dpto - Sexo	Personas
Personas afiliadas a emergencias móviles	Salud	1	Dpto - Sexo	Personas
Personas afiliadas a emergencias móviles	Salud	4	Dpto - Tramos de edad	Personas
Personas por tipo de atención en salud	Salud	-	Dpto	Personas
Tasa de desempleo	Mercado Laboral	1	Dpto - Sexo	Personas
Subempleo	Mercado Laboral	1	Dpto - Sexo	Personas
Tasa de empleo	Mercado Laboral	1	Dpto - Sexo	Personas
Tasa de actividad	Mercado Laboral	1	Dpto - Sexo	Personas
Ocupados en establecimientos fuera de su departamento	Mercado Laboral	1	Dpto - Sexo	Personas
Desocupados por última ocupación	Mercado Laboral	2	Dpto - Sexo	Personas
Ocupados por categoría de ocupación	Mercado Laboral	2	Dpto - Sexo	Personas
Tasa de actividad	Mercado Laboral	5	Dpto - Tramos de edad	Personas

APÉNDICE A. ANEXOS

Población total por condición de actividad	Mercado Laboral	-	Dpto - Sexo	Personas
Informalidad	Mercado Laboral / Ingresos y bienestar	1	Dpto - Sexo	Personas
Tasa de desempleo	Mercado Laboral / Ingresos y bienestar	5	Dpto - Tramos de edad	Personas
Tasa de empleo	Mercado Laboral / Ingresos y bienestar	5	Dpto - Tramos de edad	Personas
Personas en hogares en situación de pobreza	Ingresos y bienestar	1	Dpto - Sexo	Personas
Jóvenes entre 14 y 29 años en situación de pobreza	Ingresos y bienestar	1	Dpto - Sexo	Personas
Hogares con al menos 1 auto o camioneta	Ingresos y bienestar	8	Dpto	Hogares
Hogares en situación de pobreza	Ingresos y bienestar / Viviendas y hogares	8	Dpto	Hogares
Hogares con hacinamiento	Ingresos y bienestar / Viviendas y hogares	8	Dpto	Hogares

A.1. Listado de indicadores

Personas con celular	Tecnología y comunicación	1	Dpto - Sexo	Personas
Utilización de computadora el último mes	Tecnología y comunicación	1	Dpto - Sexo	Personas
Personas que utilizan internet	Tecnología y comunicación	1	Dpto - Sexo	Personas
Frecuencia de utilización de internet	Tecnología y comunicación	2	Dpto - Sexo	Personas
Hogares con al menos una computadora del Plan Ceibal	Tecnología y comunicación	8	Dpto	Hogares
Hogares con conexión a internet	Tecnología y comunicación	8	Dpto	Hogares
Hogares con computadora o laptop	Tecnología y comunicación	8	Dpto	Hogares
Población por lugar de nacimiento	Demografía	2	Dpto - Sexo	Personas
Personas por lugar de residencia anterior	Demografía	2	Dpto - Sexo	Personas
Población por lugar de residencia hace 5 años	Demografía	2	Dpto - Sexo	Personas
Hogares por tipo de evacuación del sistema sanitario	Viviendas y hogares	9	Dpto	Hogares
Hogares por presencia de lugar para cocinar	Viviendas y hogares	9	Dpto	Hogares
Hogares por presencia y uso de baño	Viviendas y hogares	9	Dpto	Hogares
Hogares por fuente de energía para iluminar	Viviendas y hogares	9	Dpto	Hogares

APÉNDICE A. ANEXOS

Hogares por origen de agua para beber y cocinar	Viviendas y hogares	9	Dpto	Hogares
Hogares por tipo de relación con la vivienda	Viviendas y hogares	-	Dpto	Hogares

A.2. Listado de variables

Se presenta la tabla con las variables de la ECH que son utilizadas en la app.

Tabla A.2: Listado de variables utilizadas en la app

Variables	Descripción	Base de datos
dpto	Código correlativo del 1 al 19 comenzando por Montevideo y continuando alfabéticamente	Personas
e26	Sexo	Personas
e27	Edad	Personas
e215	Asistencia a magisterio o profesorado	Personas
e218	Asistencia a universidad o similar	Personas
e221	Asistencia a terciario no universitario	Personas
e224	Asistencia a postgrado	Personas
pobpcoac	Condición de actividad económica	Personas
e193	Asistencia a educación preescolar	Personas
e197	Asistencia a educación primaria	Personas
e201	Asistencia a educación media	Personas
e212	Asistencia a educación técnica	Personas
e51_2	Años aprobados en primaria común	Personas
e51_4	Años aprobados en ciclo básico	Personas
e51_5	Años aprobados en bachillerato	Personas
e51_6	Años aprobados en bachillerato tecnológico	Personas
e51_7	Años aprobados en educación técnica	Personas

e51_8	Años aprobados en magisterio o profesorado	Personas
e51_9	Años aprobados en universidad o similar	Personas
e51_10	Años aprobados en terciario no universitario	Personas
e51_11	Años aprobados en postgrado	Personas
e51_7_1	Exigencia para realizar curso de educación técnica	Personas
e51_3	Años aprobados en primaria especial	Personas
e236	Lugar de residencia 5 años atrás	Personas
numero	Número de identificación del hogar	Personas
pesoano	Ponderador del año	Personas
e238	Asistencia a educación inicial	Personas
e48	Sabe leer y escribir	Personas
e46	Afiliación a emergencia móvil	Personas
f82	Aporte a caja de jubilaciones	Personas
f80	Trabajo en el departamento	Personas
f116	Ha trabajado antes	Personas
f121	Categoría de la ocupación (trabajos anteriores de no ocupados)	Personas
f73	Categoría de la ocupación	Personas
pobre06	Pobreza según metodología 2006	Personas
e62	Utilización de internet en el último mes	Personas
e61	Utilización de microcomputador en el último mes	Personas
e60	Tenencia de teléfono celular	Personas
e37	Lugar de residencia inmediato luego del nacimiento	Personas
e39	Lugar de residencia anterior	Personas
e45_1	ASSE	Personas
e45_2	IAMC	Personas
e45_3	Seguro médico privado	Personas
e45_4	Hospital Policial/Militar	Personas
e45_5	Área de salud del BPS	Personas

APÉNDICE A. ANEXOS

e45_6	Policlínica municipal	Personas
e45_7	Otro (salud)	Personas
e65	Frecuencia de utilización de internet	Personas
subempleo	Trabajador subempleado	Personas
d8_1	Tenencia de la vivienda	Hogares
d25	Total de personas del hogar	Hogares
d9	Habitaciones residenciales	Hogares
d21_18	Automóvil o camioneta	Hogares
d21_16	Conexión a internet	Hogares
d21_15	Microcomputador	Hogares
d21_15_1	Algún microcomputador del Plan Ceibal	Hogares
d16	Evacuación del sistema sanitario	Hogares
d18	Fuente de energía para iluminar	Hogares
d11	Origen del agua	Hogares
d19	Lugar para cocinar	Hogares
d15	Uso de baño	Hogares
upm_fic	UPM	UPM y estratos
estrato	Estratos	UPM y estratos
