Estimación en dominios de indicadores socioeconómicos a partir de la Encuesta Continua de Hogares

Estudiante:

- Mauricio Pittamiglio

Tutores:

- Ignacio Álvarez-Castro
- Juan José Goyeneche

Introducción

- Indicadores
- Encuesta Continua de Hogares (ECH)
- Métodos estadísticos
- Métodos computacionales
- Discusión

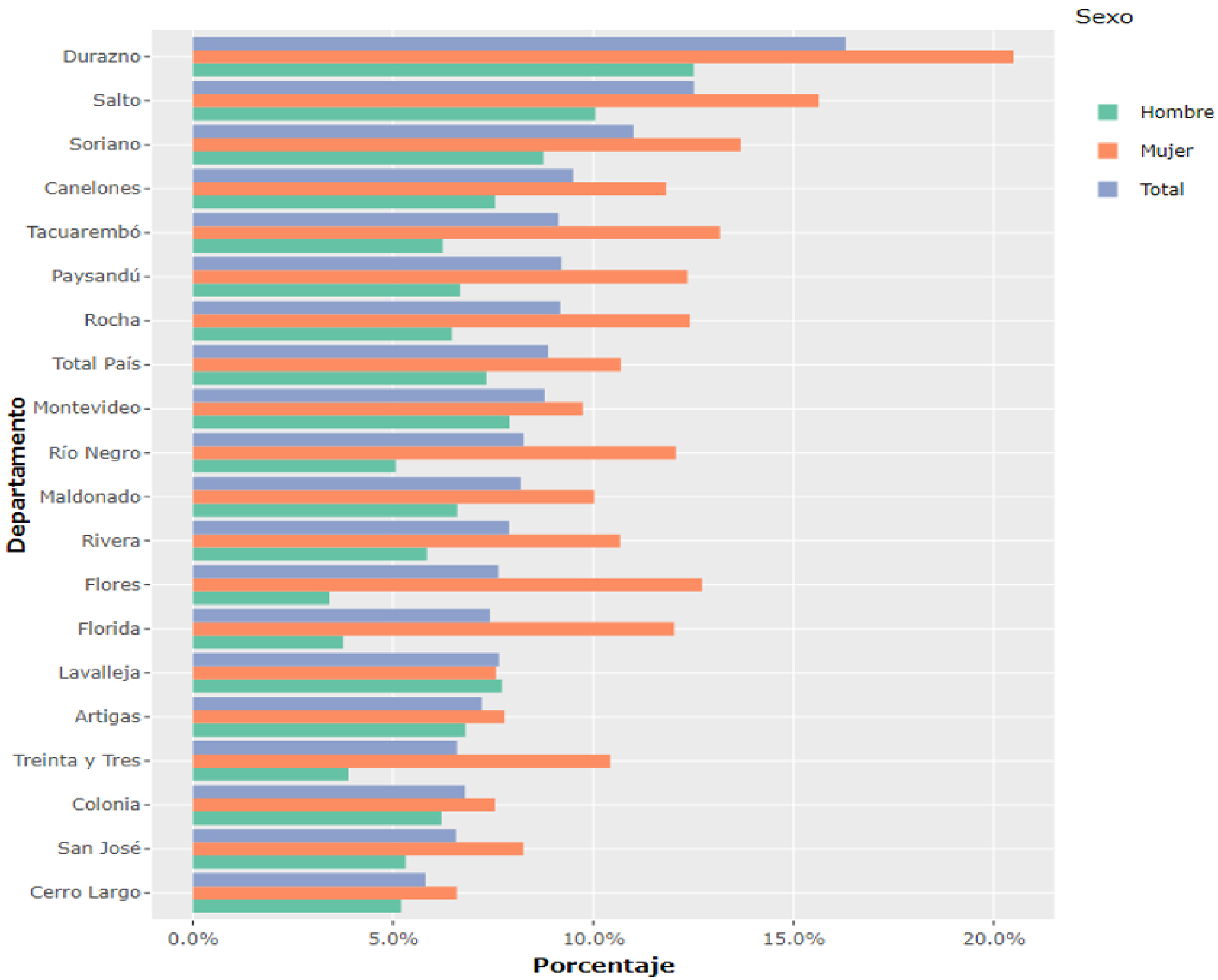
Indicadores

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).

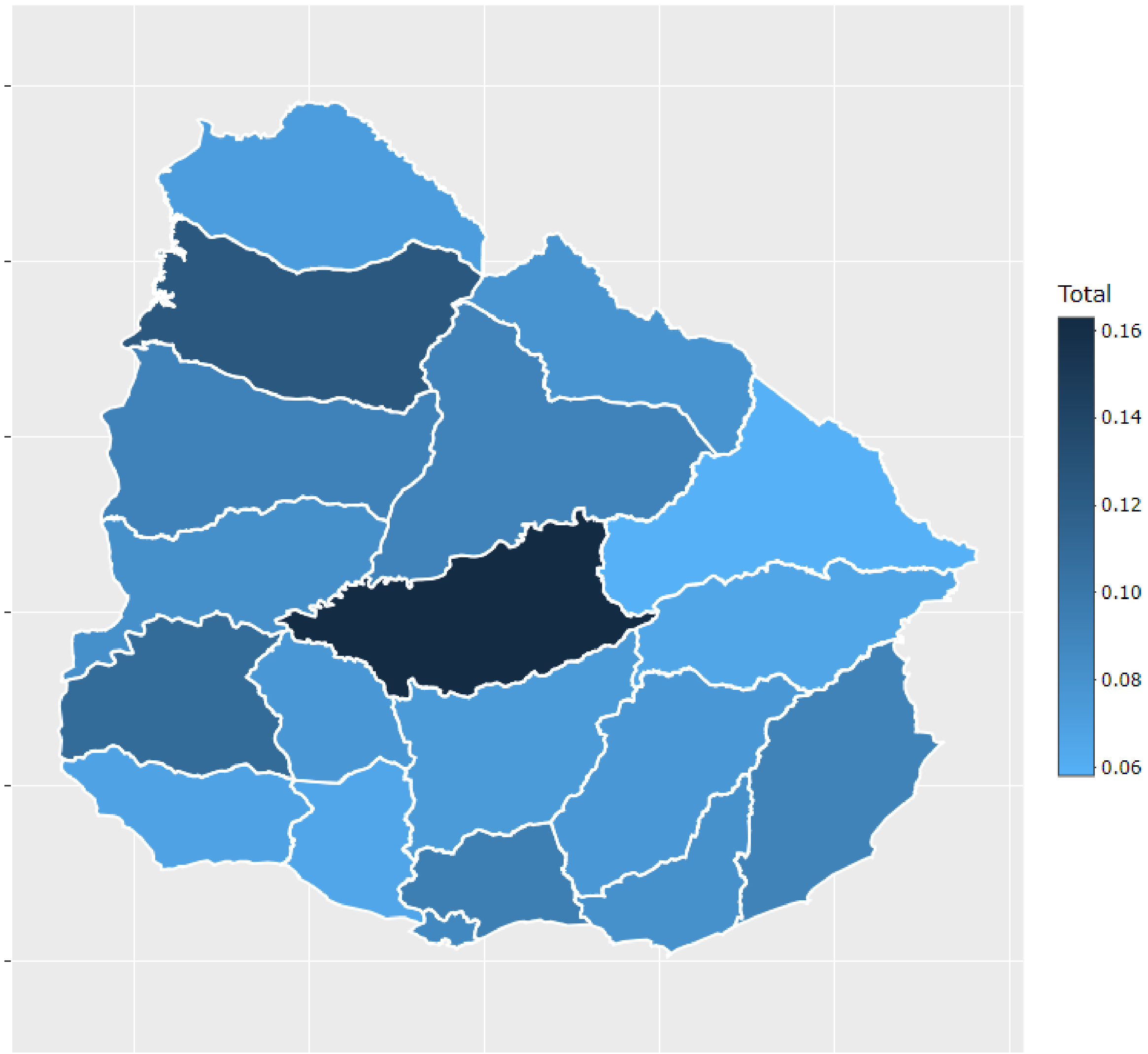
Categorías: educación, salud, mercado laboral, ingresos y bienestar, tecnología y comunicación, demografía, viviendas y hogares.

En total son 53 indicadores socio-económicos, todos desagregados por departamento y a su vez, algunos de estos se desagregan también por sexo o tramos etarios.

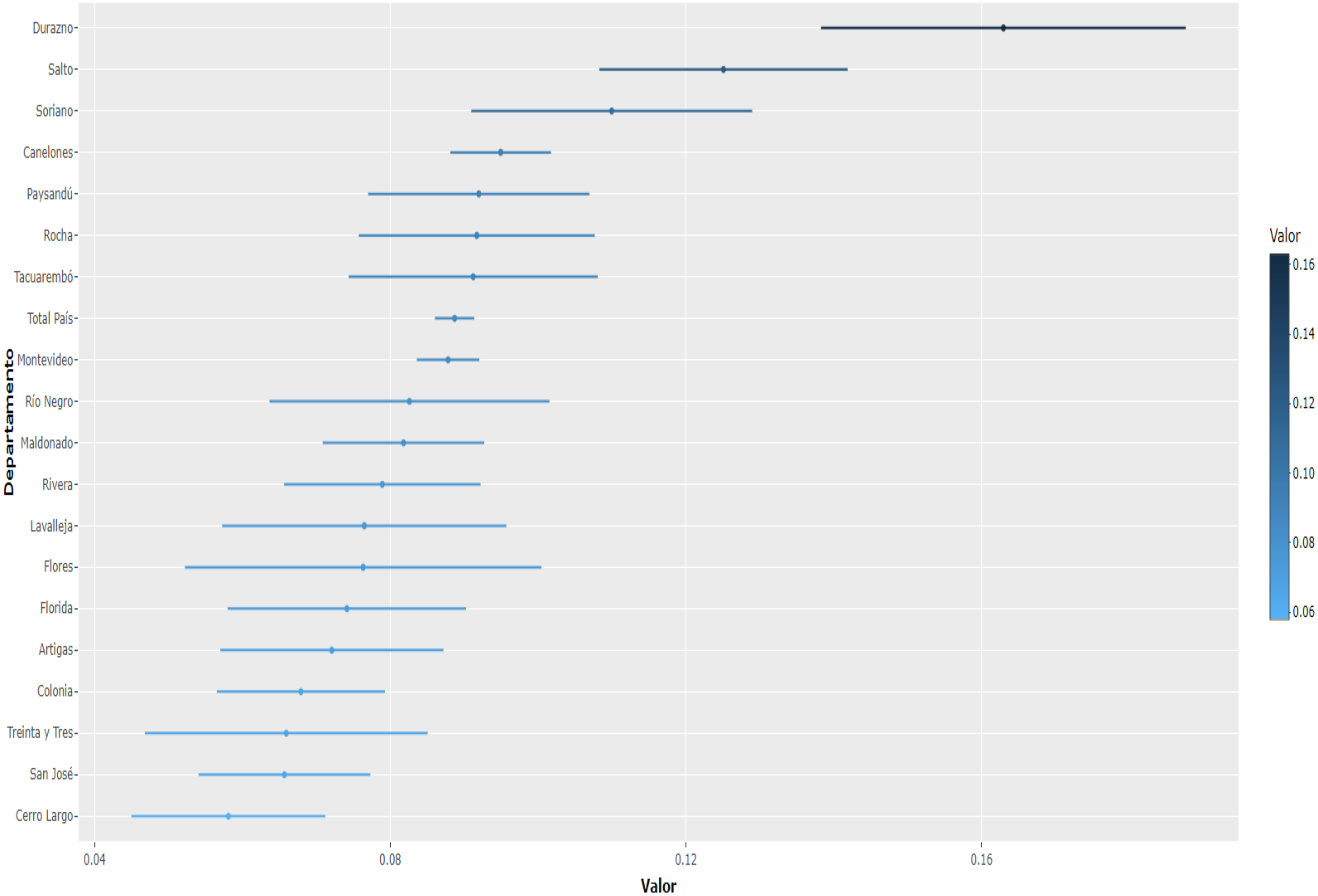
Ejemplo Tasa de desempleo



Ejemplo Tasa de desempleo



Ejemplo Tasa de desempleo



ECH

- Realizada por el INE desde 1968.
- Brinda indicadores oficiales del mercado laboral y de ingresos.

Diseño muestral

- Estratificado
- Por conglomerados
- Varias etapas de selección

Diseño muestral

- Se divide la población U en H estratos y se selecciona una muestra s_h en c/u.
- Dentro de cada estrato se encuentran M_h UPM de las cuales se seleccionan m_h con probabilidad proporcional al tamaño, en base a una medida de tamaño (MOS).
- N_j cantidad de viviendas de la UPM según censo del 2011.
- Se seleccionan n_{jh} unidades dentro de cada UPM seleccionada previamente, bajo un MAS.
- El número de unidades seleccionadas en cada UPM es fijo, es decir $n_j = \alpha$.

Métodos estadísticos

- Notación y definiciones
- Estimación puntual e IC
- Errores estándar (SE)

Notación y definiciones

Dada la población objetivo U de tamaño N, se selecciona una muestra s de tamaño n.

Parámetros de interés:

Notación y definiciones

Dada la población objetivo U de tamaño N, se selecciona una muestra s de tamaño n.

Parámetros de interés:

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

Notación y definiciones

Dada la población objetivo U de tamaño N, se selecciona una muestra s de tamaño n.

Parámetros de interés:

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

Dominios

- Subpoblaciones, para las cuales se computan de manera específica, tanto estimaciones puntuales como intervalos de confianza, son denominadas dominios (Sarndal et al., 2003).
- Pueden clasificarse en planeados y no planeados

Estimaciones puntuales

Estimador de Horvitz-Thompson para t:

- $\hat{t}_\pi = \sum_s w_i \times y_i$

Donde w_i es el peso de la unidad i incluida en la muestra. Más precisamente el ponderador o peso muestral w_i se define como el inverso de la probabilidad de inclusión π_i.

Probabilidad de inclusión

La probabilidad de inclusión en la muestra de una unidad i queda definida como:

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

Probabilidad de inclusión

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:

$$\pi_{jh} = m_h \times \frac{MOS_j}{\sum_{j=1}^{M_h} MOS_j}$$

Probabilidad de inclusión

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:

$$\pi_{jh} = m_h \times \frac{MOS_j}{\sum_{j=1}^{M_h} MOS_j}$$

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

$$\pi_{i|jh} = \frac{n_{jh}}{MOS_j}$$

Estimaciones puntuales

Los indicadores trabajados son en su mayoría ratios, por lo que el estimador queda definido de la siguiente manera:

$$\widehat{R} = \frac{\sum_{i \in s} y_i \times w_i}{\sum_{i \in s} z_i \times w_i}$$

Estimaciones puntuales

Los indicadores trabajados son en su mayoría ratios, por lo que el estimador queda definido de la siguiente manera:

$$\widehat{R} = \frac{\sum_{i \in s} y_i \times w_i}{\sum_{i \in s} z_i \times w_i}$$

- Los ponderadores son ajustados por la no respuesta.

Intervalos de confianza

- Son una medida de calidad de las estimaciones.

Intervalos de confianza

- Son una medida de calidad de las estimaciones.

$$\hat{\theta} \pm z_{1-\frac{\alpha}{2}} \times \widehat{SE}\,(\hat{\theta})$$

Intervalos de confianza

- Son una medida de calidad de las estimaciones.

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

- Se trabaja con una confianza del 95%, por lo que se sustituye $z_{1-\alpha/2}$ por 1.96.

Errores estándar (SE)

- El SE es una medida de la precisión de la estimación.
- Se busca que la estimación de los SE capte de la mejor manera posible la variabilidad de las estimaciones.

Errores estándar (SE)

- El SE es una medida de la precisión de la estimación.
- Se busca que la estimación de los SE capte de la mejor manera posible la variabilidad de las estimaciones.
- Previo a 2018 no se contaba con información clave sobre el diseño muestral.

Errores estándar (SE)

- El SE es una medida de la precisión de la estimación.
- Se busca que la estimación de los SE capte de la mejor manera posible la variabilidad de las estimaciones.
- Previo a 2018 no se contaba con información clave sobre el diseño muestral.
- No existen métodos exactos para un diseño complejo como el de la ECH.

Errores estándar (SE)

- El SE es una medida de la precisión de la estimación.
- Se busca que la estimación de los SE capte de la mejor manera posible la variabilidad de las estimaciones.
- Previo a 2018 no se contaba con información clave sobre el diseño muestral.
- No existen métodos exactos para un diseño complejo como el de la ECH.
- Aproximación de una etapa (Lumley, 2010) en conjunto con la linearización de Taylor.

Aproximación de una etapa

- Utiliza únicamente la varianza de los estimadores en la primera etapa.
- Por lo tanto, a la hora de estimar la varianza del estimador del total de una variable y, se utiliza la siguiente fórmula:

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

Aproximación de una etapa

- Utiliza únicamente la varianza de los estimadores en la primera etapa.
- Por lo tanto, a la hora de estimar la varianza del estimador del total de una variable y, se utiliza la siguiente fórmula:

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

- Siendo $\hat{t}_{jh} = \sum_{s_{jh}} w_i \times y_i$ 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.

Estimación en dominios

En el caso de la estimación en dominios se debe reemplazar la variable y, 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.

Linearización de Taylor

- La aproximación por linearización de Taylor, planteada entre otros por (Sarndal et al., 2003), se basa en expresar el estimador no lineal como función de estimadores lineales.

Linearización de Taylor

- La aproximación por linearización de Taylor, planteada entre otros por (Sarndal et al., 2003), se basa en expresar el estimador no lineal como función de estimadores lineales.

- Se utiliza el desarrollo de Taylor de primer orden para aproximar el estimador de ratio:

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

Linearización de Taylor

Posteriormente se aproxima la varianza de la estimación mediante la varianza de primer orden utilizando el método planteado anteriormente

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

Linearización de Taylor

Posteriormente se aproxima la varianza de la estimación mediante la varianza de primer orden utilizando el método planteado anteriormente

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

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

Métodos computacionales

- Se utiliza la librería *srvyr* en *R*.
- Cuenta con funciones como *survey_mean*, *survey_ratio* y *survey_total*.

Estimación de los indicadores

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))))
```

Estimación de los indicadores

Se crearon 9 funciones, las cuales necesitan de los siguientes insumos:

Estimación de los indicadores

Se crearon 9 funciones, las cuales necesitan de los siguientes insumos:

- El diseño muestral
- El filtro a aplicar (de ser necesario)
- Variable(s) de interés
- Vector con categorías de la variable de interés (en caso de ser necesario)

Estimación de los indicadores

Se crearon 9 funciones, las cuales necesitan de los siguientes insumos:

- El diseño muestral
- El filtro a aplicar (de ser necesario)
- Variable(s) de interés
- Vector con categorías de la variable de interés (en caso de ser necesario)

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

Aplicación web (Shiny)

- Shiny es una librería que permite crear fácilmente aplicaciones web interactivas a partir de R.
- Contiene al menos dos componentes: UI y server.
- Programación reactiva.
- App disponible en: GitHub