Tamaños de muestra para la razón de desocupación (trimestral) y la proporción de pobreza extrema (anual) y rotaciones

CEPAL

2024-04-23

# Introducción

En este documento calculamos el tamaño de muestra necesario para los parámetros de la razón de desocupación y la proporción de pobreza extrema. Para desocupación el tamaño de muestra es definido trimestralmente, para pobreza extrema el tamaño de muestra se define anualmente. Durante el año se realizarán diferentes módulos de la encuesta y se implementa un tamaño de muestra 5-0-0, es decir un esquema en donde se visitan las UPM’s seleccionadas durante 5 períodos consecutivos para desaparecer de la selección. Durante todos los trimestres se realiza el módulo de condiciones ocupacionales y la pobreza extrema se llevará a cabo en la primera y quinta visita del total de cinco visitas en que se realizará una UPM.

|  |
| --- |
|  |

# Tamaño de muestra para desocupación

El cálculo del tamaño de muestra se llevará a cabo inicialmente para las 6 regiones definidas: Región Central, Región Chorotega, Región Pacífica Central, Región Brunca, Región Huetar Atlántico y Región Huetar Norte. Esta metodología se emplea para asegurar que se cumpla con el tamaño mínimo de muestra en cada una de estas regiones. Posteriormente, se realizarán los cálculos por área (urbano y rural) y finalmente se obtendrá el resultado a nivel nacional, todo esto siguiendo las especificaciones establecidas por el INEC.

Los tamaños de muestras resultantes para **desocupación** son los siguientes:

|  |
| --- |
|  |

y para **pobreza extrema** son

|  |
| --- |
|  |

# Procesamiento para el calculo de los tamaños de muestras.

En la presente sección describimos los diferentes insumos, funciones y librerías empleandas para la estimación del tamaño de la muestra para desocupación y pobreza extrema.

## Tablas de parámetros

Se prepararon previamente los archivos de Excel “ParametrosTablasMuestreo\_Desempleo.xlsx” y “ParametrosTablasMuestreo\_PobrezaExtrema.xlsx”, los cuales contienen información crucial para el cálculo del tamaño de muestra. Estos archivos incluyen las siguientes variables:

* **Agregación:** Nombres de las 6 regiones, 2 áreas y a nivel nacional.
* **N:** Tamaño poblacional para cada nivel de agregación.
* **M:** Número de Unidades Primarias de Muestreo (UPM) necesarias para cada nivel de agregación.
* **P:** Razón de desocupación obtenida de la Encuesta Continua de Empleo (ECE) 2022 para cada nivel de agregación.
* **r1:** Proporción de personas mayores de 15 años en cada nivel de agregación.
* **r2:** Población económicamente activa de personas mayores de 15 años.
* **b:** Promedio de número de viviendas por hogar.
* **deff:** Factor de efecto de diseño obtenido en la ECE 2022 para cada nivel de agregación.
* **nro\_vivXUPM:** Número promedio de viviendas/hogares realizadas en cada UPM para cada nivel de agregación.
* **PM:** Pérdida de muestra debido a desactualizaciones del marco.
* **TR:** Tasa de respuesta.
* **conf:** Confiabilidad, fijada en 0.95 para este caso.

Estos insumos son esenciales para llevar a cabo los cálculos de tamaño de muestra de manera precisa y ajustada a las especificaciones requeridas por el INEC. A continuación vemos el contenido de la tabla para desocupación.

|  |
| --- |
|  |

## Funciones desarroladas.

En el siguiente apartado se describen la funciones creadas para agilizar el proceso de calculo y organización de la información, estas funciones están disponible en el archivo f\_tablaMue\_desocupacion.R

### f\_tablaMue\_desocupacion

La función f\_tablaMue\_desocupacion calcula una tabla de muestreo para estimar el parámetro de interés en una agregación específica, utilizando parámetros proporcionados en un data frame. Los argumentos de entrada son:

* df\_parametrostablas: Data frame con los parámetros de las tablas de muestreo.
* str\_region: String que especifica la agregación de interés.
* cve: Coeficiente de variación deseado para la estimación.
* m: Número de viviendas deseado por PSU.

La función retorna una tabla con los resultados del cálculo de la muestra ajustada para cada agregación.

La función hace uso principalmente de la función ss4HHSp de la librería samplesize4surveys, la cual se utiliza para estimar el tamaño de muestra necesario en estudios que involucran hogares o unidades familiares, teniendo en cuenta diversos parámetros estadísticos y de diseño de la encuesta.

### f\_kishAlloc

La función f\_kishAlloc, se utiliza para calcular el tamaño de muestra ajustado de Kish en un diseño estratificado. Este método es importante dado que permite tener en cuenta las diferencias entre los estratos de la población al determinar el tamaño de muestra necesario para obtener resultados precisos y representativos.

La función requiere como entrada el vector N\_h que contiene los tamaños de los estratos, el tamaño de muestra deseado n, y un parámetro opcional I que ajusta el efecto del tamaño de los estratos en el cálculo (por defecto es 1). Luego, realiza cálculos basados en la fórmula de Kish para obtener el tamaño de muestra ajustado para cada estrato, redondeando el resultado final.

La ecuación de Kish para calcular el tamaño de muestra ajustado en un diseño estratificado se expresa de la siguiente manera:

Donde:

* es el tamaño de muestra ajustado de Kish para cada estrato.
* es el tamaño de muestra deseado.
* es el número de estratos.
* es un parámetro opcional que ajusta el efecto del tamaño de los estratos en el cálculo (por defecto es 1).
* es el tamaño de cada estrato.
* es el tamaño total de la población.
* es la proporción de cada estrato en la población total.

### deff\_new

La función deff\_new se utiliza para calcular el factor de corrección de diseño para un nuevo tamaño de muestra en comparación con un tamaño de muestra anterior. Este factor ayuda a ajustar los errores de muestreo y a garantizar la precisión de los resultados obtenidos en estudios y encuestas.

La ecuación que describe el cálculo del factor de corrección de diseño en esta función es la siguiente:

Donde:

* es el valor que se calcula para ajustar el tamaño de muestra.
* es el nuevo tamaño de muestra para el que se quiere calcular el factor de corrección.
* es el factor de corrección de diseño utilizado previamente.
* es el tamaño de muestra anterior.

Esta ecuación tiene en cuenta cómo cambia el factor de corrección de diseño al modificar el tamaño de muestra, permitiendo así ajustar adecuadamente el diseño de la investigación para mantener la precisión de los resultados.

## Tamaño de muestra por región

Para determinar el tamaño de la muestra por región, se comienza estableciendo el coeficiente de variación esperado para cada una de las regiones. En este caso, se han fijado los siguientes valores por región:

| Regiones | cve |
| --- | --- |
| Region\_Central | 0.034 |
| Region\_Chorotega | 0.048 |
| Region\_Pacifica\_Central | 0.049 |
| Region\_Brunca | 0.046 |
| Region\_Huetar\_Atlantico | 0.050 |
| Region\_Huetar\_Norte | 0.045 |

El siguiente parámetro que se estableció fue el “Número de hogares seleccionados dentro de la UPM (m = 10)”. Finalmente, los parámetros de la función ss4HHSp están disponibles en el archivo de parámetros, donde se incluyen:

* N: Tamaño de la población.
* M: Número de conglomerados en la población.
* r: Porcentaje de personas dentro de la subpoblación de interés.
* b: Tamaño promedio del hogar (número de miembros).
* P: Valor de la proporción estimada.
* conf: Nivel de confianza estadística.

Para calcular el coeficiente de correlación intraclase (), se utiliza la siguiente ecuación:

Donde representa el número promedio de viviendas por UPM multiplicado por el tamaño promedio del hogar () y el porcentaje de personas dentro de la subpoblación de interés (). Esta información se obtuvo de la encuesta anterior. Finalmente, el valor de

Después de determinar todo los parámetros podemos hacer uso de la función f\_tablaMue\_desocupacion, de la siguiente forma

source("../00\_Funciones/f\_tablaMue\_desocupacion.R")  
library(purrr)  
vivi\_PSU <- 10 ## Numero de viviendas por PSU  
  
Regiones <-  
 c(  
 "Region\_Central",  
 "Region\_Chorotega",  
 "Region\_Pacifica\_Central",  
 "Region\_Brunca",  
 "Region\_Huetar\_Atlantico",  
 "Region\_Huetar\_Norte"  
 )  
cve\_num <- c(0.034, 0.048, 0.049, 0.046, 0.05, 0.045)  
  
df\_region <- map\_df(seq\_len(length(Regiones)), function(ii) {  
 f\_tablaMue\_desocupacion(  
 df\_parametros\_desocupacion,  
 str\_region = Regiones[ii],  
 cve = cve\_num[ii] ,  
 m = vivi\_PSU  
 ) %>% mutate(Agregado = Regiones[ii])  
})  
  
# df\_region

# Tamaño de muestra nacional

El tamaño de muestra nacional se obtiene sumando las muestras calculadas para cada región. Por lo tanto, el número de UPMs a seleccionar será igual a la suma de los UPM en la muestra de cada región, es decir, 2505 UPMs. Después de determinar el tamaño de muestra por regiones, es posible calcular las medidas de error a nivel nacional siguiendo el siguiente procedimiento.

## Cacalculo del DEFF nacional esperado

1. Seleccionar los parámetros nacionales, lo cual obtenemos mediante el siguiente filtro.

temp\_nacional <- df\_parametros\_desocupacion %>%  
 filter(Agregacion == "Nacional")

1. Calcar el DEFF esperado dado la muestra se necesita contar con

deff\_parametro <- temp\_nacional$deff

* , que se obtiene como:

HouseholdsPerPSU\_old <- temp\_nacional$nro\_vivXUPM \*  
 temp\_nacional$r1 \*  
 temp\_nacional$r2 \*  
 temp\_nacional$b

* , que se obtiene como:

HouseholdsPerPSU\_new = vivi\_PSU \*  
 temp\_nacional$b \*  
 temp\_nacional$r1 \*  
 temp\_nacional$r2

Por último, se hace uso de la función deff\_new en el siguiente código

deff\_esperado\_nacional <- deff\_new(deff\_old = deff\_parametro,  
 n\_old = HouseholdsPerPSU\_old,  
 n\_new = HouseholdsPerPSU\_new)

luego el DEFF espero es de 3.837, el cual cumple con las especificaciones dadas por el INEC.

## Calculo del cve nacional

Para calcular el coeficiente de variación (cve) y el error estándar al estimar una proporción única bajo un diseño de muestra, utilizamos la función e4p. Es fundamental tener en cuenta que el coeficiente de variación debe calcularse específicamente para la población económicamente activa.

El proceso para calcular el cve y el error estándar usando la función e4p:

1. Calcular la proporción de la población económicamente activa (PEA)

PEA <- sum(temp\_nacional$N) \*temp\_nacional$r1 \* temp\_nacional$r2

1. Número de personas en la muestra

n\_per\_muestra <- sum(df\_region$PersonsInSample)

1. La razón de desocupación obtenida de la ECE 2022

P <- temp\_nacional$P

1. Estimar el cve y el error estándar.

cve\_nnal <- samplesize4surveys::e4p(  
 N = PEA,  
 n = n\_per\_muestra,  
 P = P,  
 DEFF = deff\_esperado\_nacional,  
 conf = 0.95  
)

# Tamaño de muestra por zona urbana y rural

En esta instancia, no será necesario calcular el tamaño de la muestra, sino más bien realizar el proceso de asignación de la muestra nacional. Para ello, utilizaremos la ecuación de Kish. A continuación, detallamos el procedimiento necesario:

1. Selección de parámetros según zona: Del conjunto de parámetros disponibles, seleccionamos aquellos que corresponden a las zonas urbana y rural. También realizamos algunos cálculos adicionales que serán útiles en el proceso de asignación.
2. Identificar el tamaño de la población por cada nivel de agregación.
3. Realizar el proceso de asignación de Kish, para lo que implementamos la función f\_kishAlloc con

## Calculo del DEFF por zona

Para calcular el DEFF por zona se hace uso de la función deff\_new, así

deff\_Zona <- deff\_new(  
 deff\_old = temp\_Zona$deff,  
 n\_old = (temp\_Zona$prod \* temp\_Zona$nro\_vivXUPM),  
 n\_new = (vivi\_PSU \* temp\_Zona$prod)  
)

## Calculo del cve para la zona urbana y rural

Se debe tener los elementos para el calculo de las medidas de error por zona, por tanto se calcula el número de viviendas en la muestra

vctr\_nHouseholdsInSample <- vctr\_nPSUinSample \* vivi\_PSU

Ahora se determina el número de personas en la muestra

vctr\_nPersonsInSample <- round(vctr\_nHouseholdsInSample\*temp\_Zona$prod)

Por ultimo, se hace uso de la función e4p para el calculo del cve y el margen de error.

medidas\_error\_Zona <-  
 samplesize4surveys::e4p(  
 N = temp\_Zona$N \* temp\_Zona$r1 \* temp\_Zona$r2,  
 n = vctr\_nPersonsInSample,  
 P = temp\_Zona$P,  
 DEFF = deff\_Zona,  
 conf = 0.95  
 )

Luego de tener todos los cálculos realizados se organizan en una tabla como se muestra a continuación.

| Agregado | N | cve | HouseholdsPerPSU | DEFF | PSUinSample | HouseholdsInSample |
| --- | --- | --- | --- | --- | --- | --- |
| Nacional | 2,484,855 | 2.848088 | 10 | 3.837025 | 2,505.0000 | 25,041.000 |
| Zona\_Urbana | 1,862,283 | 3.412525 | 10 | 3.718118 | 1,545.5729 | 15,455.729 |
| Zona\_Rural | 622,572 | 2.829914 | 10 | 1.510146 | 959.4271 | 9,594.271 |
| Region\_Central | 1,643,721 | 0.034000 | 10 | 1.790000 | 762.0000 | 7,617.000 |
| Region\_Chorotega | 174,448 | 0.048000 | 10 | 1.710000 | 361.0000 | 3,610.000 |
| Region\_Pacifica\_Central | 139,180 | 0.049000 | 10 | 1.350000 | 373.0000 | 3,733.000 |
| Region\_Brunca | 148,152 | 0.046000 | 10 | 1.290000 | 381.0000 | 3,808.000 |
| Region\_Huetar\_Atlantico | 193,901 | 0.050000 | 10 | 1.510000 | 374.0000 | 3,737.000 |
| Region\_Huetar\_Norte | 185,453 | 0.045000 | 10 | 1.330000 | 254.0000 | 2,536.000 |

## Creando el intervalo de confianza

Para estimar los intervalos de confianza ejecutamos el siguiente bloque de código. En la primera parte agregamos el valor de a la tabla anterior.

df\_desocupacion <- inner\_join(df\_desocupacion,  
 df\_parametros\_desocupacion %>% select(Agregado = Agregacion, P), by = "Agregado")

Ahora, se hace uso de la función e4p para estimar los margenes de error como se muestra en la siguiente código.

tab\_desocupacion <- samplesize4surveys::e4p(  
 N = df\_desocupacion$N,  
 n = df\_desocupacion$PersonsInSample ,  
 P = df\_desocupacion$P,  
 DEFF = df\_desocupacion$DEFF,  
 conf = 0.95  
)

En el siguiente paso se construye los intervalos de confianza

df\_desocupacion$cve <- tab\_desocupacion$cve  
  
df\_desocupacion$me <- tab\_desocupacion$Margin\_of\_error / 100  
df\_desocupacion$LI\_desoc <- df\_desocupacion$P - df\_desocupacion$me  
df\_desocupacion$LS\_desoc <- df\_desocupacion$P + df\_desocupacion$me

| Agregado | P | me | LI\_desoc | LS\_desoc |
| --- | --- | --- | --- | --- |
| Nacional | 0.1195 | 0.006722383 | 0.1127776 | 0.1262224 |
| Zona\_Urbana | 0.1181 | 0.007950244 | 0.1101498 | 0.1260502 |
| Zona\_Rural | 0.1237 | 0.006956700 | 0.1167433 | 0.1306567 |
| Region\_Central | 0.1163 | 0.007741467 | 0.1085585 | 0.1240415 |
| Region\_Chorotega | 0.1330 | 0.012500963 | 0.1204990 | 0.1455010 |
| Region\_Pacifica\_Central | 0.0987 | 0.009486698 | 0.0892133 | 0.1081867 |
| Region\_Brunca | 0.1182 | 0.010636366 | 0.1075636 | 0.1288364 |
| Region\_Huetar\_Atlantico | 0.1133 | 0.011095177 | 0.1022048 | 0.1243952 |
| Region\_Huetar\_Norte | 0.1583 | 0.013948719 | 0.1443513 | 0.1722487 |

# Estimación del tamaño de muestra para la pobreza extrema

El tamaño de muestra para estimar la pobreza extrema se calcula utilizando los tamaños de muestra anteriores ajustados por un factor de , como se especifica en el diseño de panel rotativo. El proceso de calculo se describe a continuación.

1. Seleccionar las columnas de la necesarias de la tabla de de tamaños de muestras para desocupación.

df\_pobreza\_extrema <-  
 df\_desocupacion %>% select(  
 N,  
 M,  
 PSUinSample,  
 HouseholdsInSample,  
 HouseholdsInSample\_Adjusted,  
 HouseholdsPerPSU,  
 HouseholdsPerPSU\_Adjusted  
 )

1. Multiplicar el número de hogares en la muestra por .

df\_pobreza\_extrema$HouseholdsInSample <-  
 df\_pobreza\_extrema$HouseholdsInSample \* 1.6  
df\_pobreza\_extrema$HouseholdsInSample\_Adjusted <-  
 df\_pobreza\_extrema$HouseholdsInSample\_Adjusted \* 1.6

1. Estimar el DEFF, para lo cual se hace uso de la función deff\_new

vctr\_deff\_pobrezaExtrema <-  
 deff\_new(deff\_old = df\_parametros\_pobrezaExtrema$deff,  
 n\_old = df\_parametros\_pobrezaExtrema$nro\_vivXUPM,  
 vivi\_PSU)

1. Estimar las medidas de error

tab\_pobrezaExtrema <-  
 samplesize4surveys::e4p(  
 N = df\_desocupacion$N / 3.04,  
 n = df\_pobreza\_extrema$HouseholdsInSample ,  
 P = df\_parametros\_pobrezaExtrema$P,  
 DEFF = vctr\_deff\_pobrezaExtrema,  
 conf = 0.95)

1. Organizar los cálculos en una tabla

df\_pobreza\_extrema$cve\_pobrezaExtrema <- tab\_pobrezaExtrema$cve  
df\_pobreza\_extrema$me\_pobrezaExtrema <- tab\_pobrezaExtrema$Margin\_of\_error/100  
df\_pobreza\_extrema$rme\_pobrezaExtrema <- tab\_pobrezaExtrema$Relative\_Margin\_of\_error/100  
df\_pobreza\_extrema$P <- df\_parametros\_pobrezaExtrema$P  
df\_pobreza\_extrema$r <- df\_parametros\_pobrezaExtrema$r  
df\_pobreza\_extrema$b <- df\_parametros\_pobrezaExtrema$b  
df\_pobreza\_extrema$deff\_estimado <- vctr\_deff\_pobrezaExtrema  
df\_pobreza\_extrema$m <- vivi\_PSU  
df\_pobreza\_extrema$PM <- df\_parametros\_pobrezaExtrema$PM  
df\_pobreza\_extrema$PM <- df\_parametros\_pobrezaExtrema$TR  
df\_pobreza\_extrema$conf <- df\_parametros\_pobrezaExtrema$conf  
df\_pobreza\_extrema$Agregacion <- df\_parametros\_pobrezaExtrema$Agregacion  
   
  
df\_pobreza\_extrema <- df\_pobreza\_extrema %>% relocate(Agregacion)

1. Calcular los limite inferior y superior para pobreza externa

df\_pobreza\_extrema$LI\_PE <- df\_pobreza\_extrema$P - df\_pobreza\_extrema$me\_pobrezaExtrema  
df\_pobreza\_extrema$LS\_PE <- df\_pobreza\_extrema$P + df\_pobreza\_extrema$me\_pobrezaExtrema

1. Presentar los resultados.

df\_pobreza\_extrema %>%  
 select(  
 Agregado = Agregacion ,  
 N,  
 cve = cve\_pobrezaExtrema,  
 HouseholdsPerPSU,  
 HouseholdsPerPSU\_Adjusted,  
 DEFF = deff\_estimado ,  
 PSUinSample,  
 HouseholdsInSample,  
 HouseholdsInSample\_Adjusted,  
 P,  
 me = me\_pobrezaExtrema,  
 LI\_PE,  
 LS\_PE  
 )

# Asignación de los tamaños de muestras por Región - Area

El proceso de asignación que implementamos fue el Procedimiento de Ajuste Proporcional Iterativo (IPFP) el cual se utiliza para ajustar una tabla de contingencia de acuerdo con los márgenes conocidos (Número de PSUs). Esto implica modificar los valores dentro de la tabla para que las sumas marginales de las filas y columnas coincidan con los totales reales conocidos (Totales.Col y Totales.Fila). El IPFP utiliza el método de Proporción de Equilibrado para lograr este ajuste.

Al ajustar la tabla mediante el IPFP, se busca obtener una representación más precisa de la PSUs, asegurando que los totales marginales reflejen con precisión los totales reales conocidos.

## Procedimiento de Ajuste Proporcional Iterativo (IPFP)

Supongamos que tenemos una tabla de contingencia con dimensiones , donde representa las filas y representa las columnas. También tenemos los totales conocidos de las filas () y los totales conocidos de las columnas (). El objetivo es ajustar los valores de para que los totales marginales coincidan con los totales conocidos de la población.

1. **Inicialización:**
   * Se inicia con una tabla de contingencia que contiene valores observados.
   * Se tienen los totales conocidos de las filas () y los totales conocidos de las columnas ().
   * Se establece una tolerancia para determinar la convergencia del ajuste.
2. **Iteración del Procedimiento:**
   * Se inicia una iteración del IPFP.
   * Para cada celda de la tabla , se calcula el valor ajustado utilizando la ecuación de Proporción de Equilibrio:
   * Se calcula el error residual como la diferencia absoluta entre los totales marginales ajustados y los totales conocidos de la población:
   * Si es menor que la tolerancia , se considera que el ajuste ha convergido y se detiene el proceso.
   * Si es mayor que , se repite el proceso de ajuste con los nuevos valores de calculados.
3. **Convergencia:** El proceso iterativo continúa hasta que el error residual sea menor que la tolerancia , lo que indica que los totales marginales de la tabla ajustada coinciden de manera aceptable con los totales conocidos.
4. **Resultado Final:** Al finalizar las iteraciones, se obtiene una tabla ajustada que refleja con mayor precisión los totales conocidos en términos de proporciones marginales.

## Asignación de las PSU

Para la asignación de las PSUs, disponemos de la siguiente información que detalla el recuento de personas por estrato. A partir de estos datos, calculamos las proporciones de personas en cada estrato, lo que se convierte en el insumo inicial para aplicar el proceso de asignación.

| Estrato | NombreEstrato | agregacion | Zona | Nh\_region |
| --- | --- | --- | --- | --- |
| 1 | Central Bajo | Region\_Central | Urbano | 197430 |
| 2 | Central Medio | Region\_Central | Urbano | 389233 |
| 3 | Central Alto | Region\_Central | Urbano | 81840 |
| 4 | Central Rural | Region\_Central | Rural | 103976 |
| 5 | Chorotega Urbano | Region\_Chorotega | Urbano | 51319 |
| 6 | Chorotega Rural | Region\_Chorotega | Rural | 42879 |
| 7 | Pacifico Central Urbano | Region\_Pacifica\_Central | Urbano | 47044 |
| 8 | Pacifico Central Rural | Region\_Pacifica\_Central | Rural | 25737 |
| 9 | Brunca Urbano | Region\_Brunca | Urbano | 41240 |
| 10 | Brunca Rural | Region\_Brunca | Rural | 53964 |
| 11 | Huetar Caribe Urbano | Region\_Huetar\_Atlantico | Urbano | 63207 |
| 12 | Huetar Caribe Rural | Region\_Huetar\_Atlantico | Rural | 48237 |
| 13 | Huetar Norte Urbano | Region\_Huetar\_Norte | Urbano | 31581 |
| 14 | Huetar Norte Rural | Region\_Huetar\_Norte | Rural | 59294 |

En el siguiente código mostramos la tabla cruzada con las proporciones por Región - Área

TablaUrbanoRural <- matrix(  
 c(  
 0.866, 0.134,  
 0.545, 0.455,  
 0.647, 0.353,  
 0.434, 0.566,  
 0.568, 0.432,  
 0.348, 0.652  
 ), nrow = 6, ncol = 2, byrow = T  
)  
  
colnames(TablaUrbanoRural) <- c("Urbano", "Rural")  
rownames(TablaUrbanoRural) <- c(  
 "Central",  
 "Chorotega",  
 "Pacifica\_Central",  
 "Brunca",  
 "Huetar\_Atlantico",  
 "Huetar\_Norte"  
)  
  
# TablaUrbanoRural

El siguiente paso implica tener los totales marginales de las PSUs, los cuales fueron obtenidos previamente.

Col.knw <- c(1546, 959)  
Row.knw <- c(762, 361, 373, 381, 374, 254)

Ahora, haciendo uso de la función IPFP de la librería TeachingSampling se realiza el proceso de la asignación de la muestra.

library(tidyr)  
tabla\_PSU <- IPFP(TablaUrbanoRural, Col.knw, Row.knw, tol = 0.0001)  
data.frame(tabla\_PSU[1:6,1:2]) %>% round() %>%   
 tibble::rownames\_to\_column(var = "Region")

Region Urbano Rural  
1 Central 657 105  
2 Chorotega 194 167  
3 Pacifica\_Central 238 135  
4 Brunca 162 219  
5 Huetar\_Atlantico 209 165  
6 Huetar\_Norte 86 168

El paso siguiente es realizar la asignación de las viviendas como se muestra en la siguiente tabla

|  |
| --- |
|  |