

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.

Año / Mes	Mes 1					Mes 2					Mes 3				
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
2025-III	A1	B1	C1	D1	E1	F1	G1	H1	I1	J1	K1	L1	M1	N1	O1
2025-IV	A1	B1	C1	D1	E2	F1	G1	H1	I1	J2	K1	L1	M1	N1	O2
2026-I	A1	B1	C1	D2	E2	F1	G1	H1	I2	J2	K1	L1	M1	N2	O2
2026-II	A1	B1	C2	D2	E2	F1	G1	H2	I2	J2	K1	L1	M2	N2	O2
2026-III	A1	B2	C2	D2	E2	F1	G2	H2	I2	J2	K1	L2	M2	N2	O2
2026-IV	A2	B2	C2	D2	E2	F2	G2	H2	I2	J2	K2	L2	M2	N2	O2
2027-I	A2	B2	C2	D2	E3	F2	G2	H2	I2	J3	K2	L2	M2	N2	O3
2027-II	A2	B2	C2	D3	E3	F2	G2	H2	I3	J3	K2	L2	M2	N3	O3
2027-III	A2	B2	C3	D3	E3	F2	G2	H3	I3	J3	K2	L2	M3	N3	O3
2027-IV	A2	B3	C3	D3	E3	F2	G3	H3	I3	J3	K2	L3	M3	N3	O3

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:

Agregado	N	cve	HouseholdsPerPSU	HouseholdsPerPSU_Adjusted	DEFF	PSUinSample	HouseholdsInSample	HouseholdsInSample_Adjusted	P	me	LI_desoc	LS_desoc
Nacional	2484855	2,87	10	12,21	3,84	2505	25041	30824	0,1195	0,006722383	0,112777617	0,126222383
Zona_Urbana	1862283	3,43	10	12,49	3,72	1546	15456	19310	0,1181	0,007950244	0,110149756	0,126050244
Zona_Rural	622572	2,87	10	11,81	1,51	959	9594	11335	0,1237	0,0069567	0,1167433	0,1306567
Region_Central	1643721	3,40	10	12,50	1,79	762	7617	9527	0,1163	0,007741467	0,108558533	0,124041467
Region_Chorotega	174448	4,80	10	11,69	1,71	361	3610	4219	0,133	0,012500963	0,120499037	0,145500963
Region_Pacifica_Central	139180	4,90	10	12,79	1,35	373	3733	4770	0,0987	0,009486698	0,089213302	0,108186698
Region_Brunca	148152	4,59	10	12,08	1,29	381	3808	4601	0,1182	0,010636366	0,107563634	0,128836366
Region_Huetar_Atlantico	193901	5,00	10	12,49	1,51	374	3737	4673	0,1133	0,011095177	0,102204823	0,124395177
Region_Huetar_Norte	185453	4,50	10	11,94	1,33	254	2536	3034	0,1583	0,013948719	0,144351281	0,172248719

y para **pobreza extrema** son

Agregado	N	cve	HouseholdsP	HouseholdsPerPSU	HouseholdsPerPSU_Adjusted	DEFF	PSUinSample	HouseholdsInSample	HouseholdsInSample_Adjusted	P	me	LI_PE	LS_PE
Nacional	2484855	2,09	10	12,21	1,25	2505	40066	49319	0,0637996	0,00260937	0,06119024	0,06640897	
Zona_Urbana	1862283	2,99	10	12,49	1,38	1546	24729	30896	0,05641654	0,00330372	0,05311282	0,05972026	
Zona_Rural	622572	2,50	10	11,81	0,94	959	15351	18137	0,08328437	0,00408029	0,07920407	0,08736466	
Region_Central	1643721	4,84	10	12,50	1,34	762	12187	15244	0,04377859	0,0041501	0,0396285	0,04792869	
Region_Chorotega	174448	4,58	10	11,69	1,15	361	5776	6751	0,07883091	0,00708042	0,07175049	0,08591133	
Region_Pacifica_Central	139180	4,05	10	12,79	1,24	373	5973	7632	0,09917485	0,00788126	0,09129359	0,10705611	
Region_Brunca	148152	3,62	10	12,08	1,04	381	6093	7362	0,10208818	0,00724692	0,09484126	0,10933511	
Region_Huetar_Atlantico	193901	3,29	10	12,49	0,84	374	5979	7476	0,10526247	0,00677832	0,09848415	0,11204079	
Region_Huetar_Norte	185453	5,19	10	11,94	1,23	254	4058	4854	0,0948532	0,00964532	0,08520788	0,10449851	

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.

Agregacion	N	M	P	r1	r2	b	deff	nro_vivXUPM	PM	TR
Nacional	2,484,855	23,484	0.1195	0.60	0.79	3.040368	3.855316	10.060000	0.08	0.89
Zona_Urbana	1,862,283	12,798	0.1181	0.62	0.80	3.031809	3.671000	9.838178	0.08	0.87
Zona_Rural	622,572	10,686	0.1237	0.57	0.76	3.062290	1.528000	10.323601	0.08	0.92
Region_Central	1,643,721	4,626	0.1163	0.63	0.80	3.049600	1.788000	9.930393	0.07	0.86
Region_Chorotega	174,448	3,812	0.1330	0.56	0.76	3.069563	1.738000	10.326338	0.07	0.92
Region_Pacifica_Central	139,180	2,816	0.0987	0.57	0.78	2.978529	1.342000	9.848366	0.08	0.85
Region_Brunca	148,152	3,971	0.1182	0.52	0.77	2.904828	1.300000	10.156887	0.10	0.92
Region_Huetar_Atlantico	193,901	4,297	0.1133	0.55	0.75	2.997394	1.506000	9.892716	0.08	0.87
Region_Huetar_Norte	185,453	3,962	0.1583	0.57	0.75	3.167365	1.339000	10.184250	0.09	0.92

Funciones desarrolladas.

En el siguiente apartado se describen las funciones creadas para agilizar el proceso de cálculo y organización de la información, estas funciones están disponibles 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 `ss4HHS` 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:

$$n_h^{Kish} = n \cdot \frac{\sqrt{\frac{1}{H^2} + I \cdot \left(\frac{N_h}{N}\right)^2}}{\sum \sqrt{\frac{1}{H^2} + I \cdot \left(\frac{N_h}{N}\right)^2}}$$

Donde:

- n_h^{Kish} es el tamaño de muestra ajustado de Kish para cada estrato.
- n es el tamaño de muestra deseado.
- H es el número de estratos.
- I es un parámetro opcional que ajusta el efecto del tamaño de los estratos en el cálculo (por defecto es 1).
- N_h es el tamaño de cada estrato.
- N es el tamaño total de la población.
- $\frac{N_h}{N}$ 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:

$$def f_{new} = 1 + (n_{new} - 1) \times \rho_y = 1 + (n_{new} - 1) \times \frac{def f_{old} - 1}{n_{old} - 1}$$

Donde:

- $def f_{new}$ es el valor que se calcula para ajustar el tamaño de muestra.
- n_{new} es el nuevo tamaño de muestra para el que se quiere calcular el factor de corrección.
- $def f_{old}$ es el factor de corrección de diseño utilizado previamente.
- n_{old} 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 (ρ), se utiliza la siguiente ecuación:

$$\rho = \frac{(def - 1)}{(\bar{n}_{encuesta} - 1)}$$

Donde $\bar{n}_{encuesta}$ representa el número promedio de viviendas por UPM multiplicado por el tamaño promedio del hogar (b) y el porcentaje de personas dentro de la subpoblación de interés (r). Esta información se obtuvo de la encuesta anterior. Finalmente, el valor de $\delta = cve \times Z_{\alpha/2}$

Después de determinar todos 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.

Calculo del DEFF nacional esperado

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

```
temp_nacional <- df_parametros_desocupacion %>%  
  filter(Agregacion == "Nacional")
```

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

- $deff_{old}$

```
deff_parametro <- temp_nacional$deff
```

- n_{old} , que se obtiene como:

```
HouseholdsPerPSU_old <- temp_nacional$nro_vivXUPM *  
  temp_nacional$r1 *  
  temp_nacional$r2 *  
  temp_nacional$b
```

- n_{new} , 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
```

2. Número de personas en la muestra

```
n_per_muestra <- sum(df_region$PersonsInSample)
```

3. La razón de desocupación obtenida de la ECE 2022

```
P <- temp_nacional$P
```

4. 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 $I = 1$

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

```
vcctr_nHouseholdsInSample <- vcctr_nPSUinSample * vivi_PSU
```

Ahora se determina el número de personas en la muestra

```
vcctr_nPersonsInSample <- round(vcctr_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$r2,
    n = vcctr_nPersonsInSample,
    P = temp_Zona$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	25,041.0000
Zona_Urbana	1,862,283	3.412525	10	3.718118	1,545.5729	15,455.7290
Zona_Rural	622,572	2.829914	10	1.510146	959.4271	9,594.2710
Region_Central	1,643,721	0.034000	10	1.790000	762.0000	7,617.0000
Region_Chorotega	174,448	0.048000	10	1.710000	361.0000	3,610.0000
Region_Pacifica_Central	139,180	0.049000	10	1.350000	373.0000	3,733.0000
Region_Brunca	148,152	0.046000	10	1.290000	381.0000	3,808.0000
Region_Huetar_Atlantico	193,901	0.050000	10	1.510000	374.0000	3,737.0000
Region_Huetar_Norte	185,453	0.045000	10	1.330000	254.0000	2,536.0000

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 P a la tabla anterior.

```
df_desocupacion <- inner_join(df_desocupacion,
                              df_parametros_desocupacion %>% select(Agregado = Agregacion,
```

Ahora, se hace uso de la función `e4p` para estimar los márgenes 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 1.6, 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
)
```

2. Multiplicar el número de hogares en la muestra por 1.6.

```
df_pobreza_extrema$HouseholdsInSample <-
  df_pobreza_extrema$HouseholdsInSample * 1.6
df_pobreza_extrema$HouseholdsInSample_Adjusted <-
  df_pobreza_extrema$HouseholdsInSample_Adjusted * 1.6
```

3. Estimar el DEFF, para lo cual se hace uso de la función `deff_new`

```
vcctr_deff_pobrezaExtrema <-
  deff_new(deff_old = df_parametros_pobrezaExtrema$deff,
           n_old = df_parametros_pobrezaExtrema$nro_vivXUPM,
           vivi_PSU)
```

4. 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 = vcctr_deff_pobrezaExtrema,
    conf = 0.95)
```

5. 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$deff_estimado <- vcctr_deff_pobrezaExtrema
df_pobreza_extrema$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)
```

6. 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
```

7. 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 *Table* con dimensiones $m \times n$, donde m representa las filas y n representa las columnas. También tenemos los totales conocidos de las filas (*Row.knw*) y los totales conocidos de las columnas (*Col.knw*). El objetivo es ajustar

los valores de *Table* para que los totales marginales coincidan con los totales conocidos de la población.

1. Inicialización:

- Se inicia con una tabla de contingencia *Table* que contiene valores observados.
- Se tienen los totales conocidos de las filas (*Row.knw*) y los totales conocidos de las columnas (*Col.knw*).
- Se establece una tolerancia *tol* para determinar la convergencia del ajuste.

2. Iteración del Procedimiento:

- Se inicia una iteración del IPFP.
- Para cada celda (i, j) de la tabla *Table*, se calcula el valor ajustado utilizando la ecuación de Proporción de Equilibrio:

$$Table_{ij} = \frac{Col.knw_i \times Row.knw_j}{GrandTotal}$$

- Se calcula el error residual como la diferencia absoluta entre los totales marginales ajustados y los totales conocidos de la población:

$$Error_{residual} = \sum_{i=1}^m |TotalMarginal_i - TotalConocido_i|$$

- Si $Error_{residual}$ es menor que la tolerancia *tol*, se considera que el ajuste ha convergido y se detiene el proceso.
 - Si $Error_{residual}$ es mayor que *tol*, se repite el proceso de ajuste con los nuevos valores de *Table* calculados.
3. **Convergencia:** El proceso iterativo continúa hasta que el error residual sea menor que la tolerancia *tol*, 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 *Table* 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	Nombre Estrato	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

	Muestra Trimestral Fuerza de trabajo		Muestra anual Pobreza	
	Tamaño de muestra de viviendas	Tamaño de la muestra (UPM)	Tamaño de muestra de viviendas	Tamaño de la muestra (UPM)
Total	30825	2505	49320	2505
Zona				
Urbana	19310	1546	30896	1546
Rural	11515	959	18424	959
Región Central	9527	762	15243	762
Urbana	8214	657	13143	657
Bajo	2426	194	3882	194
Medio	4783	382	7653	382
Alto	1005	81	1608	81
Rural	1313	105	2100	105
Región de Chorotega	4219	361	6750	361
Urbana	2267	194	3628	194
Rural	1952	167	3123	167
Región Pacifico Central	4770	373	7632	373
Urbana	3044	238	4870	238
Rural	1726	135	2762	135
Región Brunca	4602	381	7363	381
Urbana	1957	162	3131	162
Rural	2645	219	4232	219
Región Huetar Central	4673	374	7477	374
Urbana	2611	209	4178	209
Rural	2062	165	3299	165
Región Huetar Norte	3034	254	4854	254
Urbana	1027	86	1644	86
Rural	2007	168	3211	168