

Estimación del salario anual de los trabajadores domésticos a partir de la GEIH

Se inicia cargando las librerías.

```
library(purrr)
library(dplyr)
library(tempdisagg)
library(data.table)
library(readr)
```

Se enruta a la carpeta que contiene los meses del año.

```
#setwd("C:/Users/josej/Desktop/Gobernación/Tareas/Cuentas 2019/GEIH2019")
```

A continuación se ejecuta esta función; esta función lee todos los archivos de la GEIH para un año, los archivos deben estar en una carpeta con el año como nombre, y al interior de ella, en carpetas individuales con el mes al que pertenece como nombre, así: “mes.csv”.

```
geih <- function(año,bases,variables){
  bases_anuales <- vector("list",length=length(bases))
  for(i in 1:length(bases)){
    common_path <- paste(getwd(),"/",año,sep="")
    directorios <- list.files(common_path)
    combinaciones <- CJ(directorios, bases[[i]])
    sub_ruta <- paste(año,"/",combinaciones[[1]],"/",combinaciones[[2]],sep="")
    lista_mensual <- lapply(sub_ruta, function(j) {read_csv2(j)})
    for(a in 1:12){
      colnames(lista_mensual[[a]]) <-tolower(colnames(lista_mensual[[a]]))
    }
    for(j in 1:length(lista_mensual)){
      lista_mensual[[j]]$mes <-rep(substring(substring(sub_ruta[j],6), 1,
                                             regexpr(".csv", substring(sub_ruta[j],6))-1),
                                length(lista_mensual[[j]][[1]]))
    }
    for(j in 1:length(lista_mensual)){
      lista_mensual[[j]] <-lista_mensual[[j]][,variables[[i]]]
    }
    bases_anuales[[i]] <- Reduce(function(d1, d2) rbind(d1, d2), lista_mensual)
  }
  datos_totales <- data.frame()
  datos_totales <- Reduce(function(d1, d2) merge(d1, d2, all=TRUE), bases_anuales)
  rownames(datos_totales) <- NULL

  assign(paste("GEIH","_",año,sep=""), data.frame(datos_totales), envir = .GlobalEnv)
}
```

Los argumentos de la función con entonces el año, el nombre de las bases de datos y las variables a tomar de cada base de datos. Se selecciona solamente las bases de cabecera y resto, que tienen representación departamental. La base de área tiene en gran proporción las mismas unidades muestrales, de manera que no es posible estimar el salario para el área metropolitana y para el resto de municipios; solo se puede estimar el salario departamental.

```
bases <- list("Cabecera - Ocupados.csv", "Resto - Ocupados.csv")
variables <- list(c("directorio", "secuencia_p", "orden",
                   "inglabo", "p6430", "mes", "dpto"),
                 c("directorio", "secuencia_p", "orden",
                   "inglabo", "p6430", "mes", "dpto"))

GEIH_2019 <- geih(2019, bases, variables)
```

Siguiendo la metodología de la GEIH, se selecciona el ítem “Trabajadores domésticos”, y se inspecciona la base resultante.

```
geih_domestico <- filter(GEIH_2019, p6430 %in% c(3))
glimpse(geih_domestico)
```

```
## Rows: 11,517
## Columns: 7
## $ directorio <dbl> 4804612, 4804651, 4804687, 4804775, 4804781, 4804787, 480...
## $ secuencia_p <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
## $ orden <dbl> 1, 4, 2, 6, 4, 8, 5, 1, 3, 12, 2, 2, 2, 3, 2, 5, 6, 6, 4, ...
## $ inglabo <dbl> 1261242, 1101242, 870000, 2150000, 480000, 1960000, 10000...
## $ p6430 <dbl> 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, ...
## $ mes <chr> "Enero", "Enero", "Enero", "Enero", "Enero", "Enero", "En...
## $ dpto <chr> "05", "05", "05", "08", "08", "08", "08", "08", "08", "08..."
```

Se selecciona solamente el departamento de Antioquia.

```
geih_domestico_ant <- filter(geih_domestico, dpto %in% c("05"))
```

Lo siguiente es inspeccionar si existen datos faltantes en la variable de ingreso laboral. Como son proporcionalmente muy pocos, se toma la decisión de eliminarlos. La eliminación de datos faltantes no es siempre correcta, de ser una proporción mayor es preferible que se utilice alguna técnica de imputación (Restrepo y Marín, 2012).

```
table(is.na(geih_domestico_ant$inglabo))
```

```
##
## FALSE TRUE
##    601    1
```

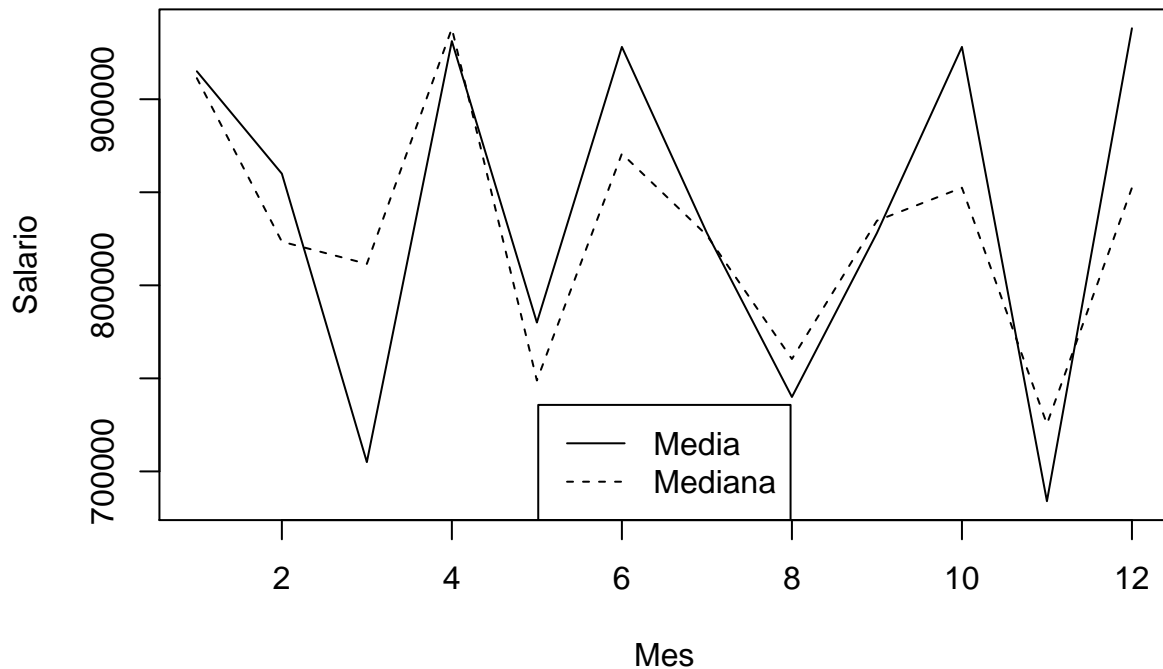
```
#Son muy pocos, se eliminan
geih_domestico_ant=geih_domestico_ant[!is.na(geih_domestico_ant$inglabo),]
#No queda ningún faltante
table(is.na(geih_domestico_ant$inglabo))
```

```
##
## FALSE
##    601
```

El salario promedio mensual no muestra el efecto estacional esperado por compensaciones como la prima, si no que muestra una variación aparentemente sin estructura. La mediana muestra divergencias importantes en algunos meses lo cual implica que la distribución tiene datos atípicos, lo cual hace preferible el uso de la mediana para reflejar correctamente el ingreso del trabajador promedio.

```
sal_med <- as.vector(tapply(X=geih_domestico_ant$inglabo,as.factor(geih_domestico_ant$mes),median))
sal_mean <- as.vector(tapply(X=geih_domestico_ant$inglabo,as.factor(geih_domestico_ant$mes),mean))

plot(sal_med,type="l",ylab="Salario",xlab="Mes")
  lines(sal_mean,lty=2)
  legend("bottom",legend=c("Media", "Mediana"),lty=1:2)
```



El salario anual se toma como la suma de las medianas de los salarios mensuales:

```
sum(sal_med)
```

```
## [1] 10065801
```

Referencias

Restrepo, M. I. y Marín, J. M. (2012). Imputación de ingresos en la Gran Encuesta Integrada de Hogares (GEIH) de 2010. Revista Desarrollo y Sociedad, (70), 219-243.