

# Muestreo por conglomerados monoetapico

Hernan Perci Nuñez Palomino

Ultima edicion 31 Mayo 2021

## Contents

<b>1</b>	<b>Introduccion</b>	<b>3</b>
<b>2</b>	<b>Diseño de la investigacion</b>	<b>3</b>
2.1	Objetivo General . . . . .	3
2.2	Objetivo Especifico . . . . .	3
2.3	Variable de estudio . . . . .	3
2.4	Poblacion objetivo . . . . .	5
2.5	Poblacion muestreada . . . . .	5
2.6	Marco Muestral . . . . .	5
2.7	Unidad de Muestreo . . . . .	5
<b>3</b>	<b>Librerias del programa R usados</b>	<b>5</b>
<b>4</b>	<b>Base de datos</b>	<b>6</b>
4.1	Tratamiento de la variable de trabajo . . . . .	10
<b>5</b>	<b>Tecnica PPT cr para los conglomerados</b>	<b>13</b>
5.1	Analisis de la variable auxiliar . . . . .	13
5.2	Tamaño de muestra mediante el coeficiente de variacion . . . . .	13
5.3	Metodo de lahiri . . . . .	16
5.4	Total estimado . . . . .	17
5.5	Media muestral por elemento . . . . .	17
5.6	Estimacion de la varianza de la media muestral por elemento . . . . .	17
5.7	Estimacion de la varianza del total estimado . . . . .	18
5.8	Seleccion de las muestras . . . . .	18
5.9	Resultados PPTcr . . . . .	18
5.10	Intervalo de confianza . . . . .	18
<b>6</b>	<b>Tecnica MASsr para los conglomerados</b>	<b>20</b>
6.1	Cuasivarianza poblacional . . . . .	20
6.2	Tamaño de muestra mediante el coeficiente de variacion . . . . .	20
6.3	Total estimado MASsr . . . . .	21
6.4	Media estimada por elementos MASsr . . . . .	21
6.5	Estimada de la varianza de la media por elemento estimado . . . . .	23
6.6	Estimacion de la varianza del total estimado . . . . .	23
6.7	Seleccion de las muestras . . . . .	23
6.8	Resultados MASsr . . . . .	23
6.9	Intervalo de confianza . . . . .	24
<b>7</b>	<b>Tecnica de media de medias con el MASsr para conglomerados</b>	<b>25</b>
7.1	Cuasivarianza poblacional de y promedio . . . . .	25

7.2	Total estimado . . . . .	25
7.3	Media estimada por elementos . . . . .	26
7.4	Estimada de la varianza de la media por elemento estimado . . . . .	26
7.5	Estimada de la varianza del total estimado . . . . .	26
7.6	Analisis del sesgo del estimador . . . . .	26
7.7	tamanio de muestra mediante el coeficiente de variacion . . . . .	27
7.8	Seleccion de las muestras . . . . .	28
7.9	Resultados Media de medias . . . . .	29
7.10	Intervalo de confianza . . . . .	29
<b>8</b>	<b>Tecnica de razon para los conglomerados</b>	<b>30</b>
8.1	Cuasivarianza de R estimado . . . . .	30
8.2	Total estimado . . . . .	31
8.3	Media estimada por elementos . . . . .	31
8.4	Estimada de la varianza de la media por elemento estimado . . . . .	31
8.5	Estimada de la varianza del total estimado . . . . .	32
8.6	tamanio de muestra mediante el coeficiente de variacion . . . . .	32
8.7	Seleccion de las muestras . . . . .	35
8.8	Resultados Razon . . . . .	35
8.9	Analisis del sesgo del estimador . . . . .	36
8.10	Intervalo de confianza . . . . .	36
<b>9</b>	<b>Conclusiones</b>	<b>38</b>

## List of Figures

1	pastos naturales . . . . .	4
2	Provincias de la region de Apurimac . . . . .	5
3	Grafico de la seleccion de la data de trabajo . . . . .	8
4	Grafico de densidad de hectareas de pastos naturales . . . . .	9
5	Grafico de densidad de hectareas de pastos naturales por distrito . . . . .	11
6	Grafico de densidad de registros de pastos naturales por distrito . . . . .	12
7	Diagrama de dispersion y recta de regresion . . . . .	14
8	Grafico del tamano de muestra en el PPTcr . . . . .	16
9	Grafico del tamano de muestra en el MASsr . . . . .	22
10	Grafico del tamano de muestra en Media de medias . . . . .	28
11	Grafico del cv(x) respecto al tamano de muestra . . . . .	34
12	Grafico del tamano de muestra en Razon . . . . .	35
13	Grafico del tamano de muestras segun tecnicas . . . . .	39
14	Grafico del coef_var estimado segun tecnicas . . . . .	40

## List of Tables

1	Caracteristicas de la variable de estudio . . . . .	5
2	Hectareas de pastos naturales en la region de Apurímac (6 primeros registros) . . . . .	6
3	Hectareas de pastos naturales en la region de Apurímac por conglomerado (6 primeros registros)	10
4	Tabla de resultados del analisis de coeficientes . . . . .	13
5	Tamaño de muestra PPTcr con respecto al coeficiente de variacion . . . . .	15
8	Tamaño de muestra MASsr con respecto al coeficiente de variacion . . . . .	21
11	Tamaño de muestra MediaMed con respecto al coeficiente de variacion . . . . .	27
14	Tamaño de muestra Razon con respecto al coeficiente de variacion . . . . .	33

# **1 Introducción**

Mediante el Decreto de Urgencia N° 055-2011 del 14 de octubre de 2011, se declara de interés y de prioridad nacional el levantamiento del IV Censo Nacional Agropecuario (IV CENAGRO) y se encarga al Instituto Nacional de Estadística e Informática (INEI) su ejecución. El Ministerio de Agricultura y el Instituto Nacional de Estadística e Informática efectuarán de manera conjunta su planeamiento, organización y dirección.

El IV CENAGRO, constituye una de las fuentes de información estadística más importantes, dado que proporciona datos actualizados para el mejor conocimiento de la estructura agraria del país. La conformación de la estructura agraria obtenida mediante el IV CENAGRO, conjuntamente con la que mide la dinámica productiva, permitirá la formulación de planes de desarrollo y la adopción de políticas que coadyuven a mejorar el nivel de vida de la población vinculada a las actividades agropecuarias.

En ese marco, se ha elaborado el Programa Censal, que es un documento de carácter técnico cuyo contenido está relacionado con los temas y variables a investigarse; las mismas que se han seleccionado en función a los requerimientos y propuestas de los usuarios públicos y privados. Cabe indicar, que la temática censal que contiene el documento es concordante con las variables que requiere el Sector Agrario para la formulación de las políticas, planes y programas de desarrollo.

## **2 Diseño de la investigación**

El uso de la tierra es el modo en que se aprovecha los terrenos de la unidad agropecuaria. La información sobre el uso de la tierra se tomará por parcelas e incluye la superficie total de la unidad agropecuaria, efectivamente aprovechada o no para la producción. La tierra se clasifica según se indica a continuación:

1. Superficie Agrícola
  - Tierras de labranza
  - Tierras dedicadas a cultivos permanentes
2. Tierras con pastos naturales
3. Tierras con montes y bosques
4. Tierras dedicadas a otros usos

Las tierras con pastos naturales son las tierras cubiertas por pastos que han crecido de modo natural, pueden ser manejadas por el hombre y estar utilizadas o no para el pastoreo de ganado.

### **2.1 Objetivo General**

Estimar las principales características de las superficies con pastos naturales de la región de Apurímac, durante el CENAGRO -2012 aplicando el muestreo de conglomerado monoetápico. Comparar los resultados de las estimaciones encontradas a través de distintas técnicas de muestreo de conglomerado monoetápico.

### **2.2 Objetivo Específico**

1. Estimar la superficie total cubierta con pastos naturales mediante las distintas técnicas de muestreo.
2. Estimar la superficie promedio cubierta con pastos naturales mediante las distintas técnicas de muestreo.
3. Comparar las técnicas de muestreo monoetápico.

### **2.3 Variable de estudio**



Figure 1: pastos naturales



Figure 2: Provincias de la region de Apurimac

Table 1: Caracteristicas de la variable de estudio

Variable	Definicion	Tipo
Y: Área de las tierras con pastos naturales(has)	Área de las tierras con pastos naturales que le pertenece al productor agropecuario en el año 2012.	Cuantitativa continua (hectareas)

## 2.4 Poblacion objetivo

Los productores agropecuarios ubicados en la región de Apurímac para el CENAGRO 2012.

## 2.5 Poblacion muestrada

Los 9125 productores agropecuarios ubicados en las 78 de los 80 distritos en la región de Apurímac para el CENAGRO 2012.

## 2.6 Marco Muestral

Los 9125 productores agropecuarios ubicados en las 78 de los 80 distritos en la región de Apurímac para el CENAGRO 2012.

## 2.7 Unidad de Muestreo

Cada uno de los distritos en la región de Apurímac para el CENAGRO 2012.

## 3 Librerias del programa R usados

```
library(tidyverse)
## -- Attaching packages ---- tidyverse 1.3.0 --
## v ggplot2 3.3.2     v purrr    0.3.4
## v tibble   3.0.3     v dplyr    1.0.1
5
```

```
library(openxlsx)
```

## 4 Base de datos

```
base_de_datos <- read_sav("01_IVCENAGRO_REC01.sav")

base_de_datos$ubigeo_distrito <- paste(base_de_datos$P001,
                                         base_de_datos$P002,
                                         base_de_datos$P003, sep = "")

data_elemento <- tibble(Distrito = base_de_datos$ubigeo_distrito,
                        Yij = base_de_datos$WSUP14)

data_elemento <- data_elemento %>%
  filter(Yij > 0)
summary(data_elemento)

##      Distrito          Yij
##  Length:18720    Min.   : 0.00
##  Class :character 1st Qu.: 0.07
##  Mode  :character Median : 0.25
##                                Mean   : 44.83
##                                3rd Qu.: 0.60
##                                Max.   :34744.32
```

Se seleccionan los registros comprendidos entre el primer y tercer cuartil.

```
data_elemento <- data_elemento %>%  
  filter(Yij > 0.07, Yij < 0.6)
```

Se convierte la variable provincia de **character** a un formato **factor** para usarlo como conglomerados.

```
data_elemento$Distrito <- as_factor(data_elemento$Distrito)
glimpse(data_elemento)
```

```

## Rows: 9,125
## Columns: 2
## $ Distrito <fct> 030101, 030101, 030101, 030101, 030101, 030101, 030101, 03...
## $ Yij      <dbl> 0.20, 0.25, 0.10, 0.59, 0.55, 0.50, 0.09, 0.50, 0.45, 0.22...
kable(head(data_elemento),
      align = "cc",
      caption = "Hectareas de pastos naturales
en la region de Apurímac (6 primeros registros)")

```

Table 2: Hectareas de pastos naturales en la region de Apurímac (6 primeros registros)

Distrito	Yij
030101	0.20
030101	0.25
030101	0.10
030101	0.59
030101	0.55
030101	0.50

```

ggplot(base_de_datos, aes(x = WSUP14)) +
  geom_density() +
  geom_rug(sides = "bl") +
  annotate(
    "text",
    x = 0.04, y = 0.02,
    label = "0.07",
    vjust = 1, size = 4, color = "red"
  ) +
  annotate(
    "text",
    x = 0.04, y = 0.2,
    label = "25%",
    vjust = 1, size = 4, color = "blue"
  ) +
  annotate(
    "text",
    x = 0.2, y = 0.2,
    label = "50%",
    vjust = 1, size = 8, color = "red"
  ) +
  annotate(
    "text",
    x = 1, y = 0.2,
    label = "25%",
    vjust = 1, size = 4, color = "blue"
  ) +
  annotate(
    "text",
    x = 1, y = 0.02,
    label = "0.6",
    vjust = 1, size = 4, color = "red"
  ) +
  geom_vline(xintercept = 0.07, color = "red") +
  geom_vline(xintercept = 0.6, color = "red") +
  labs(x = "Pastos naturales por unidad agropecuaria(has)") +
  scale_x_log10() +
  theme_stata()

```

```

## Warning: Transformation introduced infinite values in continuous x-axis
## Warning: Transformation introduced infinite values in continuous x-axis
## Warning: Removed 65556 rows containing non-finite values (stat_density).

```

Para que cumpla con la variable auxiliar Mi solo se tomo en cuenta los registros comprendidos entre el primer y tercer cuartil

```

ggplot(data_elemento, aes(x = Yij)) +
  geom_density() +
  geom_rug(sides = "bl") +
  labs(x = "Pastos naturales por unidad agropecuaria(has)") +
  theme_stata()

```

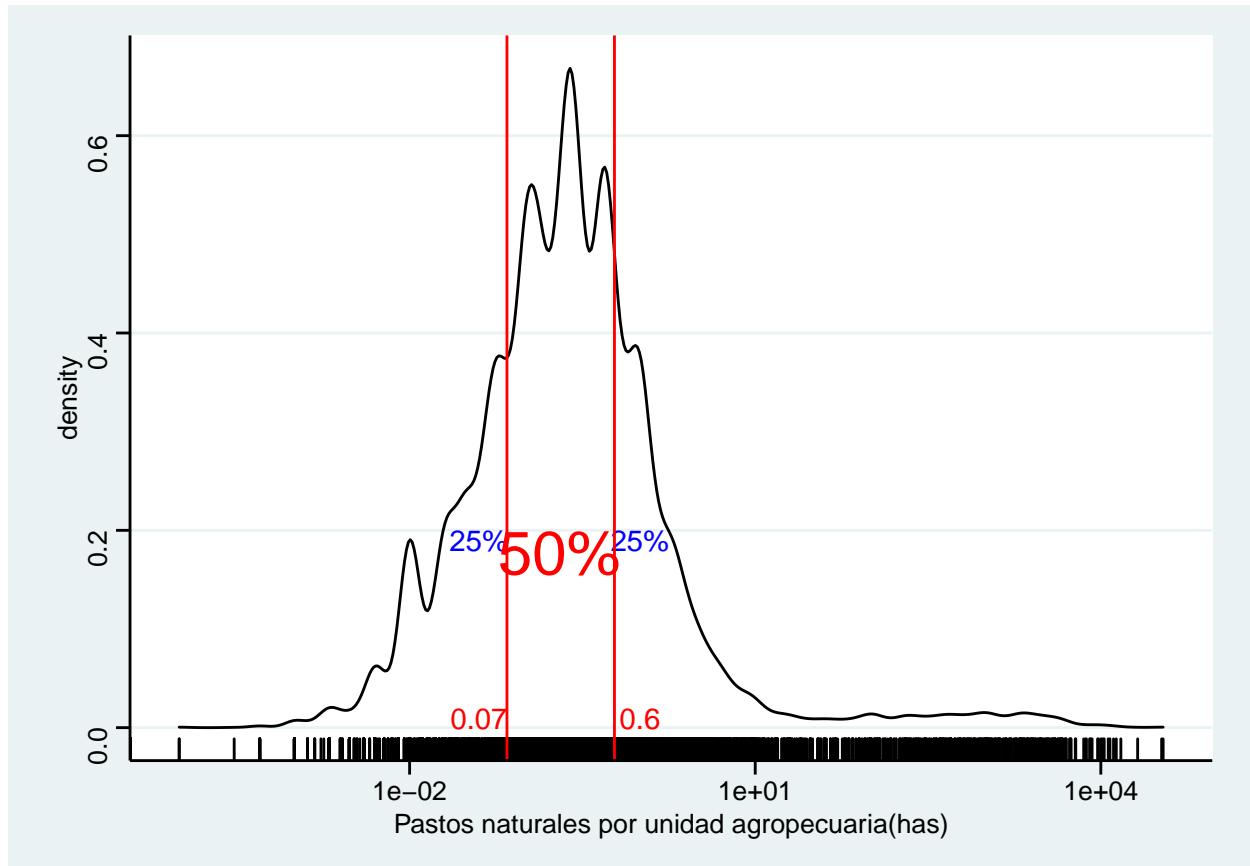


Figure 3: Grafico de la seleccion de la data de trabajo

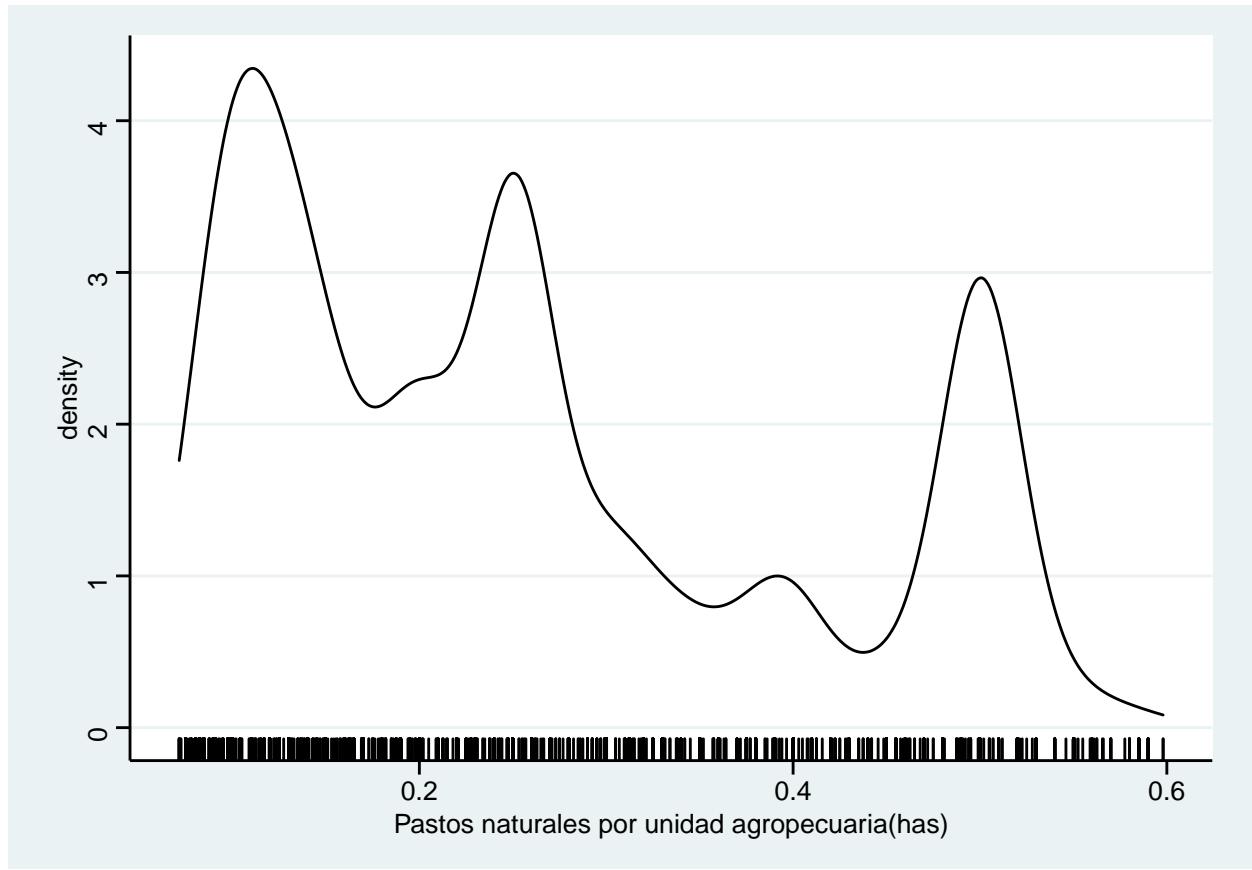


Figure 4: Grafico de densidad de hectareas de pastos naturales

## 4.1 Tratamiento de la variable de trabajo

Agrupamos por distrito para representar la data a nivel de conglomerado.

```
data_conglomerado <-
  data_elemento %>%
  group_by(Distrito) %>%
  mutate(Mi = n(), Yi = sum(Yij), Yi_prom = mean(Yij)) %>%
  select(-Yij) %>%
  unique()

data_conglomerado <-
  data_conglomerado %>%
  mutate(pi = Mi / sum(data_conglomerado$Mi),
        zi = Yi / pi)

kable(head(data_conglomerado),
      align = "ccc",
      caption = "Hectareas de pastos naturales  
en la region de Apurímac por conglomerado (6 primeros registros)")
```

Table 3: Hectareas de pastos naturales en la region de Apurímac por conglomerado (6 primeros registros)

Distrito	Mi	Yi	Yi_prom	pi	zi
030101	340	86.2460	0.2536647	0.0372603	2314.690
030102	2	0.7500	0.3750000	0.0002192	3421.875
030103	16	4.2100	0.2631250	0.0017534	2401.016
030104	192	46.0807	0.2400036	0.0210411	2190.033
030105	230	72.4800	0.3151304	0.0252055	2875.565
030106	64	10.6126	0.1658219	0.0070137	1513.125

```
ggplot(data_conglomerado, aes(x = Yi)) +
  geom_density() +
  geom_rug(sides = "bl") +
  labs(x = "Pastos naturales por unidad agropecuaria(has)") +
  theme_stata()

ggplot(data_conglomerado, aes(x = Mi)) +
  geom_density() +
  geom_rug(sides = "bl") +
  labs(x = "Numero de registros de pastos naturales por distrito") +
  theme_stata()

bases_de_datos <- function(variables) {

  base_de_datos_2 <-
  base_de_datos %>%
  filter(WSUP14 > 0.07, WSUP14 < 0.6)

  write.table(base_de_datos_2,
              file="base de datos final.csv", sep = ";", row.names = FALSE)

  write.table(data_elemento,
```

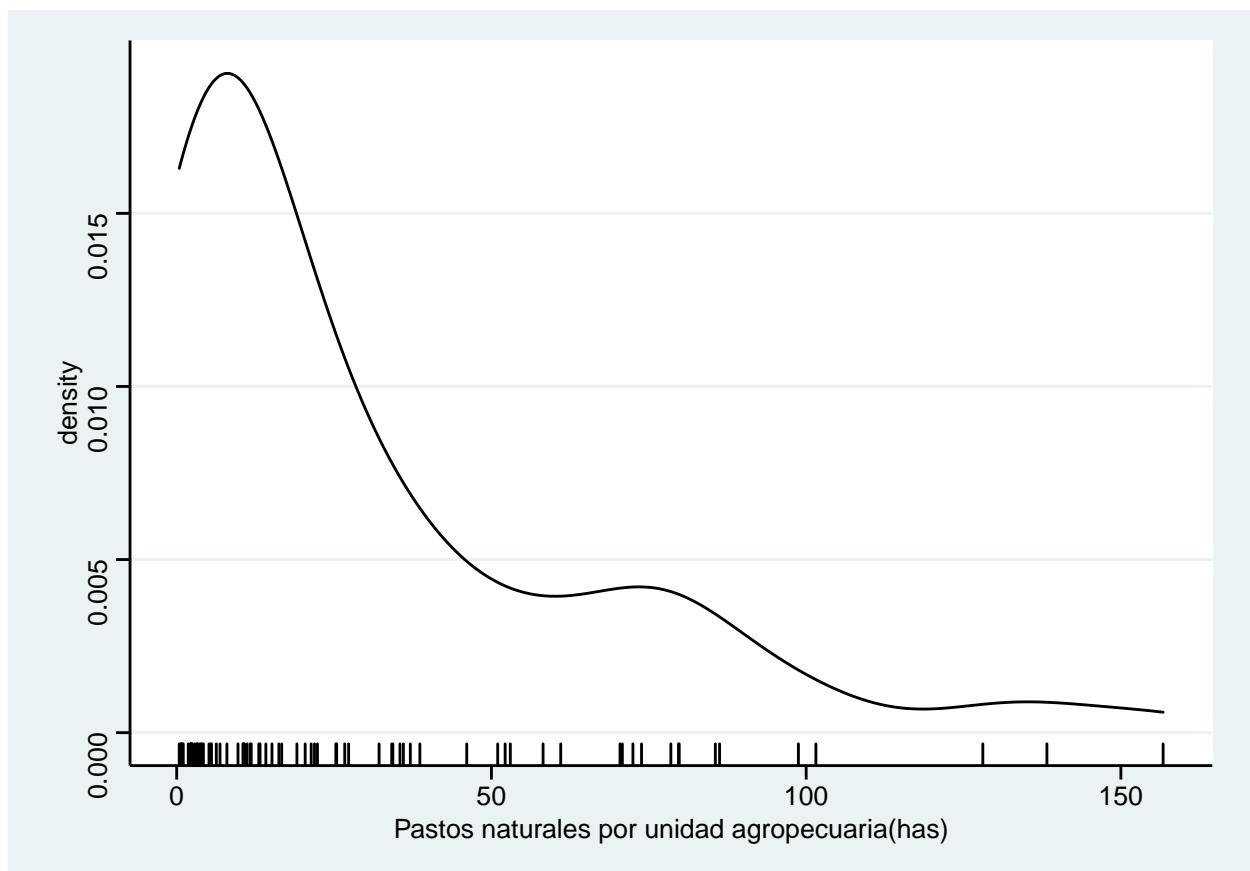


Figure 5: Grafico de densidad de hectareas de pastos naturales por distrito

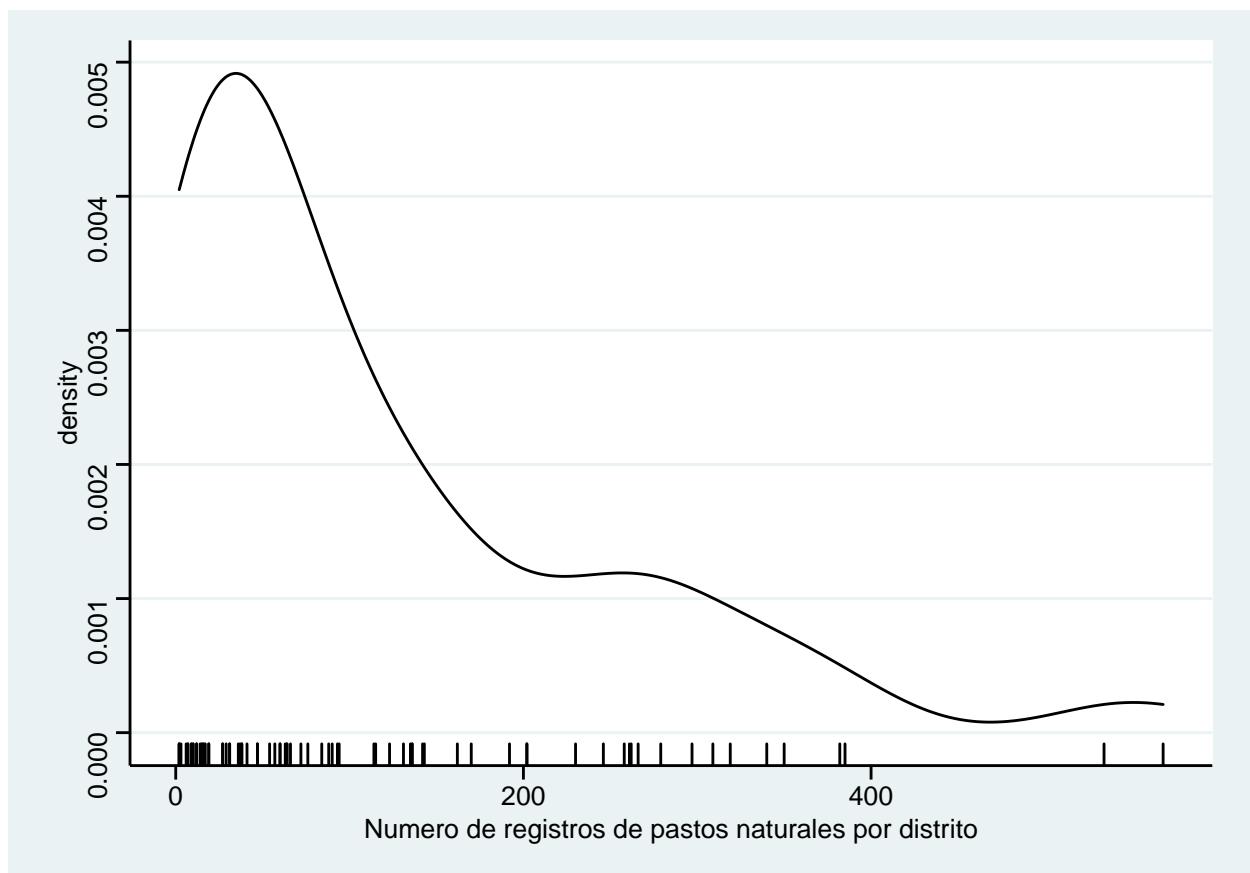


Figure 6: Grafico de densidad de registros de pastos naturales por distrito

```

    file="data por elemento.csv", sep = ";", row.names = FALSE)

write.table(data_conglomerado,
            file="data por conglomerado.csv", sep = ";", row.names = FALSE)

}

bases_de_datos()

```

## 5 Tecnica PPT cr para los conglomerados

### 5.1 Analisis de la variable auxiliar

$$H_0 : \beta_0 = 0$$

$$H_1 : \beta_0 \neq 0$$

```

modelo_lineal <- lm(formula = Yi ~ Mi,
                     data = data_conglomerado)

kable(summary(modelo_lineal)$coef,
      digits = 4,
      align = "cccc",
      caption = "Tabla de resultados del analisis
de coeficientes")

```

Table 4: Tabla de resultados del analisis de coeficientes

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-1.7493	0.8834	-1.9802	0.0513
Mi	0.2742	0.0051	53.5081	0.0000

Con un p\\_valor mayor al nivel de significancia ( $0.0513 > 0.05$ ) no hay evidencia suficiente para rechazar la hipotesis nula.

```

ggplot(data_conglomerado, aes(y = Yi, x = Mi)) +
  geom_jitter(alpha = 0.3) +
  geom_smooth(method = "lm") +
  labs(x = "Mi = variable auxiliar",
       y = "Yi = Total de Pastos naturales por distrito(has)") +
  theme_stata()

## `geom_smooth()` using formula 'y ~ x'

```

### 5.2 Tamaño de muestra mediante el coeficiente de variacion

$$n \geq \frac{M_0 \sum_{i=1}^N M_i \left( \frac{Y_i}{M_i} - \frac{Y}{M_0} \right)^2}{(a_0 Y)^2}$$

```

coef_var <- function() {
  a0 <- c(NULL)
  k <- 0.01
}

```

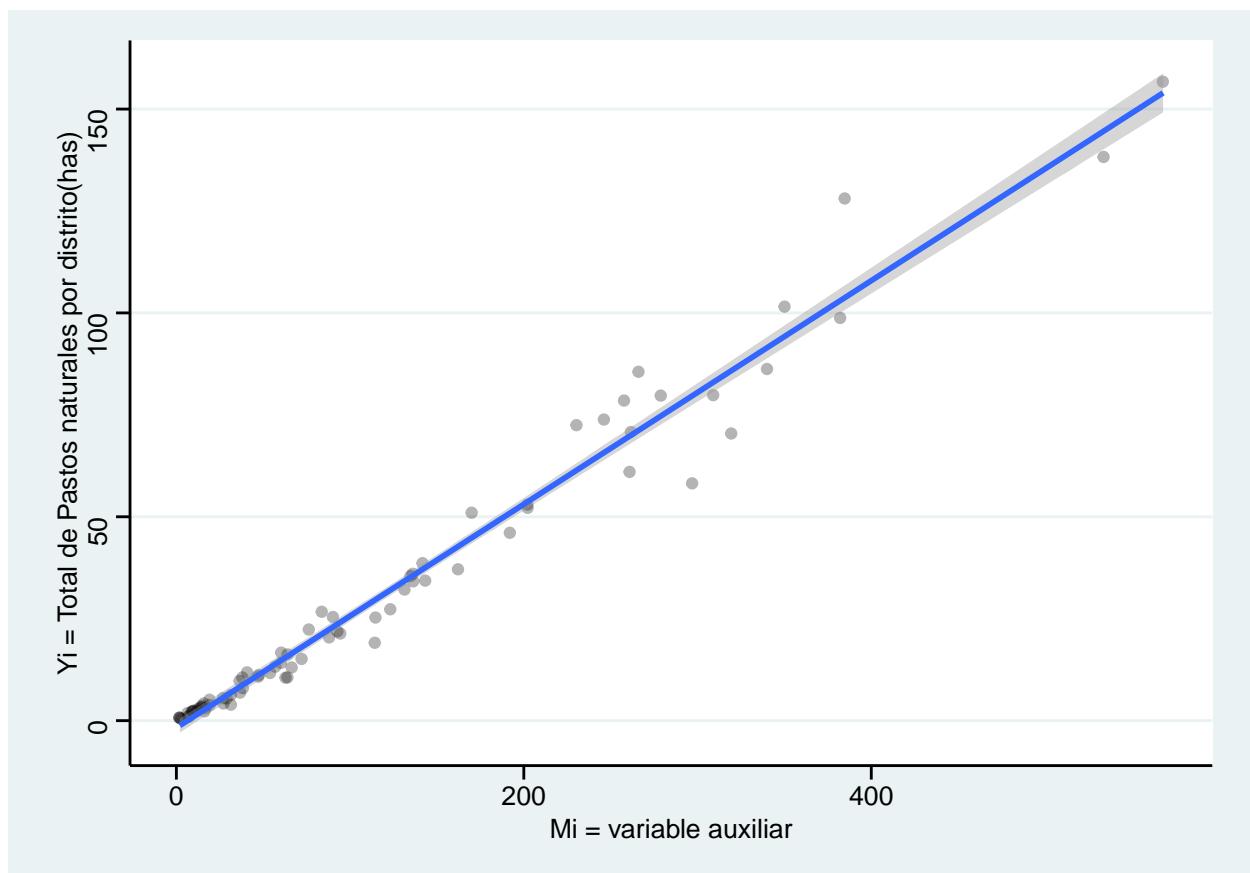


Figure 7: Diagrama de dispersion y recta de regresion

```

for (i in 1:20) {
  a0[i] <- k
  k <- k + 0.01
}
a0
}

a0 <- coef_var()

tamano_de_muestra_ppt_cr <- function(coeficiente) {
  sumatoria <- c(NULL)
  for (i in 1:nrow(data_conglomerado)) {
    sumatoria[i] <-
      data_conglomerado$Mi[i] *
        ((data_conglomerado$Yi[i] /
          data_conglomerado$Mi[i]) -
         (sum(data_conglomerado$Yi) /
           sum(data_conglomerado$Mi)))^2
  }
  numerador <- sum(data_conglomerado$Mi) * sum(sumatoria)
  denominador <- (coeficiente * sum(data_conglomerado$Yi))^2
  n <- numerador / denominador
  tabla <- tibble(a0 = coeficiente, n, n_PPTcr = ceiling(n))
  tabla
}

tabla_tamano_de_muestra_ppt_cr <- tamano_de_muestra_ppt_cr(coeficiente = a0)
kable(tabla_tamano_de_muestra_ppt_cr,
      align = "ccc",
      caption = "Tamaño de muestra PPTcr con respecto al coeficiente de variacion")

```

Table 5: Tamaño de muestra PPTcr con respecto al coeficiente de variacion

a0	n	n_PPTcr
0.01	228.0006697	229
0.02	57.0001674	58
0.03	25.3334077	26
0.04	14.2500419	15
0.05	9.1200268	10
0.06	6.3333519	7
0.07	4.6530749	5
0.08	3.5625105	4
0.09	2.8148231	3
0.10	2.2800067	3
0.11	1.8843031	2
0.12	1.5833380	2
0.13	1.3491164	2
0.14	1.1632687	2
0.15	1.0133363	2
0.16	0.8906276	1
0.17	0.7889297	1
0.18	0.7037058	1
0.19	0.6315808	1

a0	n	n_PPTcr
0.20	0.5700017	1

```
ggplot(tabla_tamano_de_muestra_ppt_cr, aes(x = a0, y = n_PPTcr)) +
  geom_label(aes(label = n_PPTcr), colour = "blue", fontface = "bold") +
  geom_line() +
  annotate(
    "text",
    x = 0.1, y = 78,
    label = "78 = Numero de conglomerados en la poblacion",
    vjust = 1, size = 3, color = "red"
  ) +
  geom_hline(yintercept = 78, color = "red") +
  labs(x = "a0 = coeficiente de variacion",
       y = "n = Tamano de la muestra") +
  theme_stata()
```

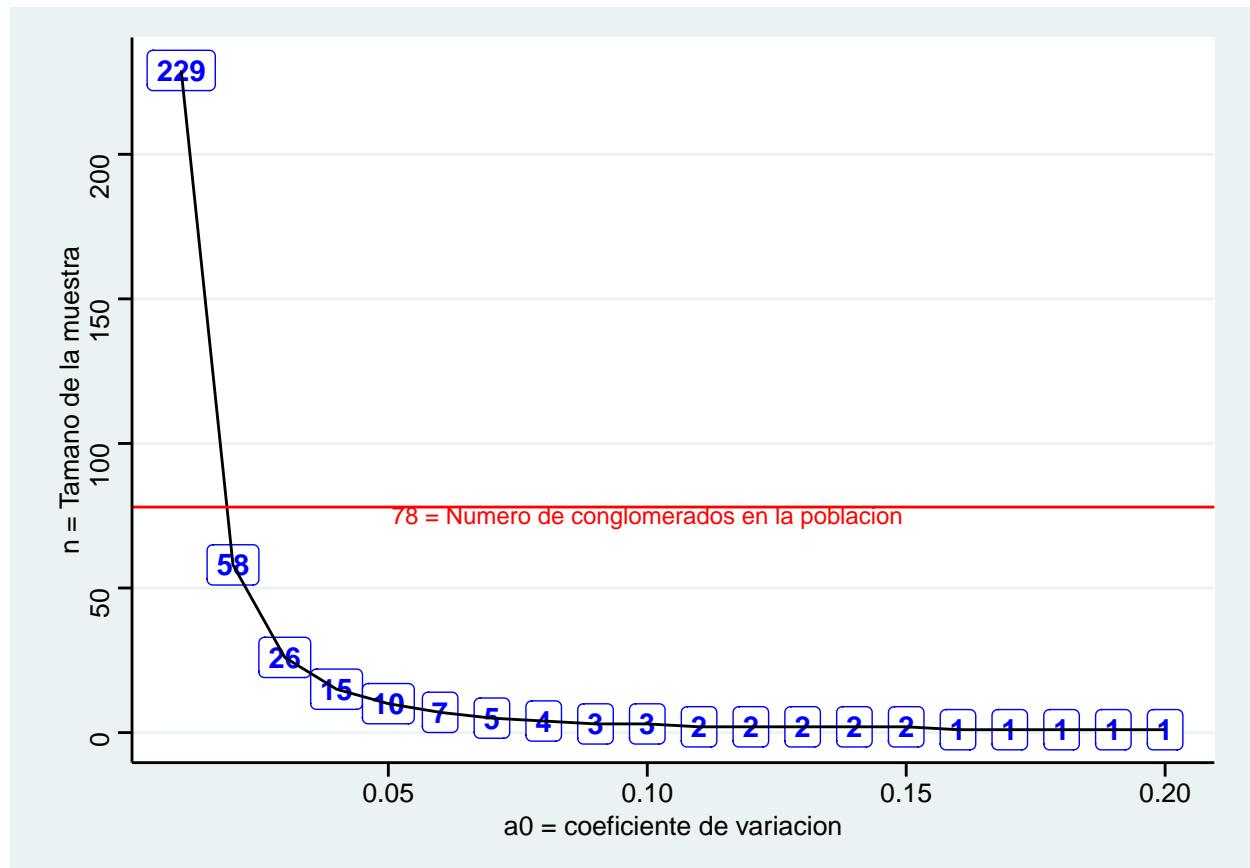


Figure 8: Grafico del tamano de muestra en el PPTcr

### 5.3 Método de Lahiri

# Lahiri, D. B. (1951). A method of sample selection providing unbiased ratio estimates  
# Bulletin of the International Statistical Institute, 33: 133 - 140.

```
# Original code from Sharon Lohr (1999), p. 452 - 453.
```

```
seleccion_lahiri_ppt_cr <- function(relsize, n, clnames = seq(along = relsize)){
  set.seed(123)
  maxrel <- max(relsize)
  sizeratio <- maxrel / mean(relsize)
  numpsu <- length(relsize)
  size <- 0
  clusters <- NULL
  while (size < n){
    ss <- ceiling((n - size) * sizeratio)
    temp <- sample(1:numpsu, ss, replace = TRUE)
    temp1 <- clnames[temp[relsize[temp] > runif(ss, min = 0, max = maxrel)]]
    clusters <- append(clusters, temp1[!is.na(temp1)])
    size <- length(clusters)
  }
  clusters[1:n]
}
```

## 5.4 Total estimado

$$\hat{Y} = \frac{M_0}{n} S \left( \frac{y_i}{M_i} \right)$$

```
total_est_PPTcr <- function(data, n = nrow(data)) {
  sumatoria <- c(NULL)
  for (i in 1:n) {
    sumatoria[i] <- data$Yi[i] / data$pi[i]
  }
  total_est <- sum(sumatoria) / n
  total_est
}
```

## 5.5 Media muestral por elemento

$$\hat{\bar{Y}}_e = \frac{\hat{Y}}{M_0}$$

```
med_mues_ele_PPTcr <- function(data) {
  media_muestral_por_elemento <- total_est_PPTcr(data) / sum(data_conglomerado$Mi)
}
```

## 5.6 Estimacion de la varianza de la media muestral por elemento

$$V(\hat{\bar{Y}}_e) = \frac{1}{M_0^2 n} \frac{S(z_i - \bar{z})^2}{n-1}$$

```
est_var_med_mues_el_pptcr <- function(data, n = nrow(data)) {
  promedio <- mean(data$zi)
  sumatoria <- c(NULL)
  for (i in 1:n) {
    sumatoria[i] <- (data$zi[i] - promedio)^2
  }
}
```

```

  est_var <- sum(sumatoria) / ((sum(data_conglomerado$Mi)^2) * n * (n - 1))
  est_var
}

```

## 5.7 Estimacion de la varianza del total estimado

$$\hat{V}(\hat{Y}) = \frac{1}{n} \frac{S(z_i - \bar{z})^2}{n-1}$$

```

est_var_total_est_PPTcr <- function(data) {
  est_var <- (sum(data_conglomerado$Mi)^2) *
    est_var_med_mues_el_pptcr(data, n = nrow(data))
  est_var
}

```

## 5.8 Seleccion de las muestras

```

data_muestra_PPTcr <- function(n) {
  indice_PPTcr <-
  seleccion_lahiri_ppt_cr(relsize =
  data_conglomerado$Mi,
  n = n)

  data_PPTcr <- data_conglomerado[indice_PPTcr,]

}

```

## 5.9 Resultados PPTcr

```

data_10_PPTcr <- data_muestra_PPTcr(n = 10)
data_3_PPTcr <- data_muestra_PPTcr(n = 3)
data_2_PPTcr <- data_muestra_PPTcr(n = 2)

estimaciones_PPTcr <- function(data) {
  tibble(total_est = total_est_PPTcr(data),
         media_ele_est = med_mues_ele_PPTcr(data),
         var_media_est = est_var_med_mues_el_pptcr(data),
         var_total_est = est_var_total_est_PPTcr(data))
}

estimaciones_data_10_PPTcr <- estimaciones_PPTcr(data = data_10_PPTcr)
estimaciones_data_3_PPTcr <- estimaciones_PPTcr(data = data_3_PPTcr)
estimaciones_data_2_PPTcr <- estimaciones_PPTcr(data = data_2_PPTcr)

```

## 5.10 Intervalo de confianza

$$\hat{Y}_e \pm t_{(1-\alpha/2;n-1)} \sqrt{\hat{V}(\hat{Y}_e)}$$

```

tabla_IC_media_PPTcr <-
  tibble(a0 = c(0.05, 0.10, 0.15),
        n = c(10, 3, 2),
        m0 = c(sum(data_10_PPTcr$Mi),

```

```

    sum(data_3_PPTcr$Mi),
    sum(data_2_PPTcr$Mi)),
media_ele = c(mean(data_elemento$Yij),
             mean(data_elemento$Yij),
             mean(data_elemento$Yij)),
media_ele_est = as.numeric(c(estimaciones_data_10_PPTcr[2],
                           estimaciones_data_3_PPTcr[2],
                           estimaciones_data_2_PPTcr[2])),
desv_media_est = sqrt(as.numeric(c(estimaciones_data_10_PPTcr[3],
                                    estimaciones_data_3_PPTcr[3],
                                    estimaciones_data_2_PPTcr[3])))

tabla_IC_media_PPTcr <-
  tabla_IC_media_PPTcr %>%
  mutate(IC_inferior = media_ele_est -
    (qt(p = 1-(0.05/2), df = n-1) * desv_media_est),
   IC_superior = media_ele_est +
    (qt(p = 1-(0.05/2), df = n-1) * desv_media_est),
   coef_var_est = desv_media_est / media_ele_est)

kable(tabla_IC_media_PPTcr)

```

a0	n	m0	media_ele	media_ele_est	desv_media_est	IC_inferior	IC_superior	coef_var_est
0.05	10	2075	0.2591973	0.2503129	0.0117541	0.2237233	0.2769025	0.0469576
0.10	3	478	0.2591973	0.2316928	0.0214080	0.1395816	0.3238039	0.0923982
0.15	2	265	0.2591973	0.2470598	0.0249867	-0.0704257	0.5645454	0.1011360

$$\hat{Y} \pm t_{(1-\alpha/2;n-1)} \sqrt{\hat{V}(\hat{Y})}$$

```

tabla_IC_total_PPTcr <-
  tibble(a0 = c(0.05, 0.10, 0.15),
  n = c(10, 3, 2),
  m0 = c(sum(data_10_PPTcr$Mi),
         sum(data_3_PPTcr$Mi),
         sum(data_2_PPTcr$Mi)),
  total = c(sum(data_conglomerado$Yi),
            sum(data_conglomerado$Yi),
            sum(data_conglomerado$Yi)),
  total_est = as.numeric(c(estimaciones_data_10_PPTcr[1],
                           estimaciones_data_3_PPTcr[1],
                           estimaciones_data_2_PPTcr[1])),
  desv_total_est = sqrt(as.numeric(c(estimaciones_data_10_PPTcr[4],
                                    estimaciones_data_3_PPTcr[4],
                                    estimaciones_data_2_PPTcr[4]))))

tabla_IC_total_PPTcr <-
  tabla_IC_total_PPTcr %>%
  mutate(IC_inferior = total_est -
    (qt(p = 1-(0.05/2), df = n-1) * desv_total_est),
   IC_superior = total_est +
    (qt(p = 1-(0.05/2), df = n-1) * desv_total_est))

```

```

  (qt(p = 1-(0.05/2), df = n-1) * desv_total_est),
  coef_var_est = desv_total_est / total_est)

kable(tabla_IC_total_PPTcr)

```

a0	n	m0	total	total_est	desv_total_est	IC_inferior	IC_superior	coef_var_est
0.05	10	2075	2365.175	2284.105	107.2560	2041.4754	2526.735	0.0469576
0.10	3	478	2365.175	2114.196	195.3480	1273.6820	2954.711	0.0923982
0.15	2	265	2365.175	2254.421	228.0032	-642.6347	5151.476	0.1011360

## 6 Técnica MASsr para los conglomerados

### 6.1 Cuasivarianza poblacional

$$S_y^2 = \frac{1}{N-1} * \sum_{i=1}^N \left( M_i * \bar{Y}_i - \sum_{i=1}^N \frac{M_i * \bar{Y}_i}{N} \right)^2$$

```

cuasivarianza_MASsr <- function(data) {
  promedio <- mean(data$Yi)
  sumatoria <- c(NULL)
  for (i in 1:nrow(data)) {
    sumatoria[i] <- (data$Yi[i] -
      promedio)^2
  }
  cuasivarianza <- sum(sumatoria) / (nrow(data) - 1)
  cuasivarianza
}

```

### 6.2 Tamaño de muestra mediante el coeficiente de variación

$$n \geq \frac{N^2 S_y^2}{a_0^2 * Y^2 + N S_y^2}$$

```

tamano_de_muestra_MASsr <- function(coeficiente) {
  CuasiPob <- cuasivarianza_MASsr(data_conglomerado)
  numerador <- (nrow(data_conglomerado)^2) * CuasiPob
  denominador <- ((coeficiente^2)*(sum(data_conglomerado$Yi)^2)) +
    (nrow(data_conglomerado)*CuasiPob)
  n <- numerador / denominador
  tabla <- tibble(a0 = coeficiente, n, n_MASsr = ceiling(n))
  tabla
}

tabla_tamano_de_muestra_MASsr <- tamano_de_muestra_MASsr(coeficiente = a0)
kable(tabla_tamano_de_muestra_MASsr,
      align = "ccc",
      caption = "Tamaño de muestra MASsr con respecto al coeficiente de variación")

```

Table 8: Tamaño de muestra MASsr con respecto al coeficiente de variacion

a0	n	n_MASsr
0.01	77.55638	78
0.02	76.25530	77
0.03	74.18120	75
0.04	71.46006	72
0.05	68.24159	69
0.06	64.68106	65
0.07	60.92436	61
0.08	57.09789	58
0.09	53.30368	54
0.10	49.61857	50
0.11	46.09628	47
0.12	42.77093	43
0.13	39.66102	40
0.14	36.77329	37
0.15	34.10608	35
0.16	31.65199	32
0.17	29.40004	30
0.18	27.33720	28
0.19	25.44951	26
0.20	23.72285	24

```
ggplot(tabla_tamano_de_muestra_MASsr, aes(x = a0, y = n_MASsr)) +
  geom_label(aes(label = n_MASsr), colour = "blue", fontface = "bold") +
  geom_line() +
  geom_hline(yintercept = 78, color = "red") +
  annotate(
    "text",
    x = 0.1, y = 78,
    label = "78 = Numero de conglomerados en la poblacion",
    vjust = 1, size = 3, color = "red"
  ) +
  labs(x = "a0 = coeficiente de variacion",
       y = "n = Tamano de la muestra") +
  theme_stata()
```

### 6.3 Total estimado MASsr

$$\hat{Y} = \frac{N}{n} \sum_{i=1}^n y_i$$

```
total_est_MASsr <- function(data) {

  total_est <- mean(data$Yi) * nrow(data_conglomerado)
  total_est
}
```

### 6.4 Media estimada por elementos MASsr

$$\hat{Y}_e = \frac{\hat{Y}}{M_0}$$

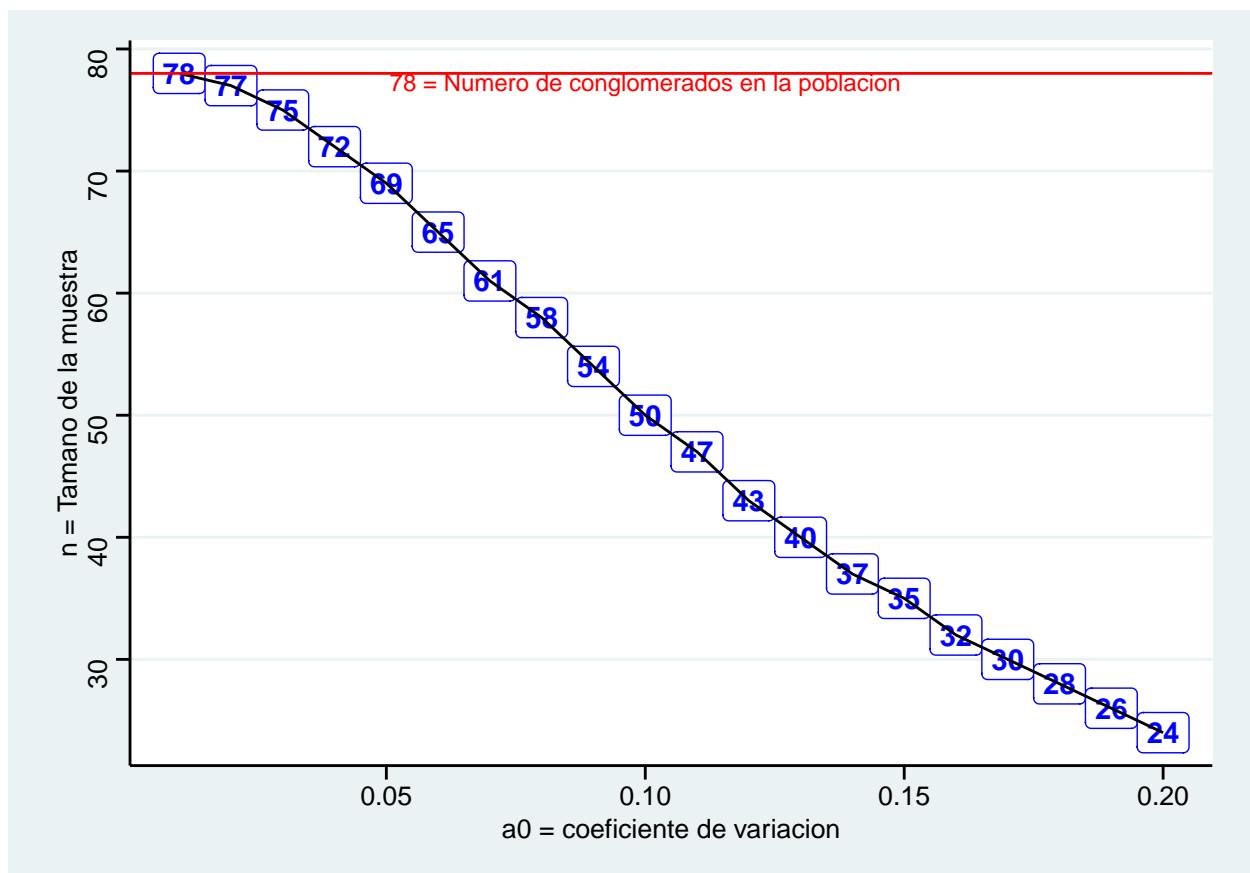


Figure 9: Grafico del tamano de muestra en el MASsr

```

med_mues_ele_MASsr <- function(data) {
  media_muestral_por_elemento <- total_est_MASsr(data) / sum(data_conglomerado$Mi)
}

```

## 6.5 Estimada de la varianza de la media por elemento estimado

$$V(\hat{Y}_e) = \frac{N^2}{M_0^2 n} \left(1 - \frac{n}{N}\right) \frac{S(y_i - \bar{y})^2}{n-1}$$

```

est_var_med_mues_el_MASsr <- function(data, n = nrow(data), N = nrow(data_conglomerado)) {

  a <- (N^2) / (n * (sum(data_conglomerado$Mi)^2))
  b <- 1 - (n / N)
  c <- cuasivarianza_MASsr(data)
  est_var <- a * b * c
  est_var

}

```

## 6.6 Estimacion de la varianza del total estimado

$$V(\hat{Y}) = M_0^2 * V(\hat{Y}_e)$$

```

est_var_total_est_MASsr <- function(data) {
  est_var <- (sum(data_conglomerado$Mi)^2) *
    est_var_med_mues_el_MASsr(data, n = nrow(data))
  est_var
}

```

## 6.7 Seleccion de las muestras

```

data_muestra_MASsr <- function(n, N = nrow(data_conglomerado)) {

  set.seed(123)
  indice_MASsr <- sample(1:N, size = n, replace = FALSE)
  data_MASsr <- data_conglomerado[indice_MASsr, 1:3]

  data_MASsr
}

```

## 6.8 Resultados MASsr

```

data_69_MASsr <- data_muestra_MASsr(n = 69)
data_50_MASsr <- data_muestra_MASsr(n = 50)
data_35_MASsr <- data_muestra_MASsr(n = 35)

estimaciones_MASsr <- function(data) {
  tibble(total_est = total_est_MASsr(data),
        media_ele_est = med_mues_ele_MASsr(data),
        var_media_est = est_var_med_mues_el_MASsr(data),
}

```

```

        var_total_est= est_var_total_est_MASsr(data))
}

estimaciones_data_69_MASsr <- estimaciones_MASsr(data = data_69_MASsr)
estimaciones_data_50_MASsr <- estimaciones_MASsr(data = data_50_MASsr)
estimaciones_data_35_MASsr <- estimaciones_MASsr(data = data_35_MASsr)

```

## 6.9 Intervalo de confianza

$$\hat{Y}_e \pm t_{(1-\alpha/2;n-1)} \sqrt{\hat{V}(\hat{Y}_e)}$$

```

tabla_IC_media_MASsr <-
  tibble(a0 = c(0.05, 0.10, 0.15),
  n = c(69, 50, 35),
  m0 = c(sum(data_69_MASsr$Mi),
         sum(data_50_MASsr$Mi),
         sum(data_35_MASsr$Mi)),
  media_ele = c(mean(data_elemento$Yij),
                mean(data_elemento$Yij),
                mean(data_elemento$Yij)),
  media_ele_est = as.numeric(c(estimaciones_data_69_MASsr[2],
                               estimaciones_data_50_MASsr[2],
                               estimaciones_data_35_MASsr[2])),
  desv_media_est = sqrt(as.numeric(c(estimaciones_data_69_MASsr[3],
                                    estimaciones_data_50_MASsr[3],
                                    estimaciones_data_35_MASsr[3]))))

tabla_IC_media_MASsr <-
  tabla_IC_media_MASsr %>%
  mutate(IC_inferior = media_ele_est -
    (qnorm(p = 1-(0.05/2)) * desv_media_est),
  IC_superior = media_ele_est +
    (qnorm(p = 1-(0.05/2)) * desv_media_est),
  coef_var_est = desv_media_est / media_ele_est)

kable(tabla_IC_media_MASsr)

```

a0	n	m0	media_ele	media_ele_est	desv_media_est	IC_inferior	IC_superior	coef_var_est
0.05	69	7811	0.2591973	0.2498878	0.0117781	0.2268031	0.2729725	0.0471337
0.10	50	5348	0.2591973	0.2356052	0.0236189	0.1893130	0.2818974	0.1002477
0.15	35	3541	0.2591973	0.2205598	0.0356349	0.1507168	0.2904028	0.1615655

$$\hat{Y} \pm t_{(1-\alpha/2;n-1)} \sqrt{\hat{V}(\hat{Y})}$$

```

tabla_IC_total_MASsr <-
  tibble(a0 = c(0.05, 0.10, 0.15),
  n = c(69, 50, 35),
  m0 = c(sum(data_69_MASsr$Mi),
         sum(data_50_MASsr$Mi),

```

```

        sum(data_35_MASsr$Mi)),
total = c(sum(data_conglomerado$Yi),
        sum(data_conglomerado$Yi),
        sum(data_conglomerado$Yi)),
total_est = as.numeric(c(estimaciones_data_69_MASsr[1],
                        estimaciones_data_50_MASsr[1],
                        estimaciones_data_35_MASsr[1])),
desv_total_est = sqrt(as.numeric(c(estimaciones_data_69_MASsr[4],
                                    estimaciones_data_50_MASsr[4],
                                    estimaciones_data_35_MASsr[4]))))

tabla_IC_total_MASsr <-
  tabla_IC_total_MASsr %>%
  mutate(IC_inferior = total_est -
        (qnorm(p = 1-(0.05/2)) * desv_total_est),
  IC_superior = total_est +
        (qnorm(p = 1-(0.05/2)) * desv_total_est),
  coef_var_est = desv_total_est / total_est)

kable(tabla_IC_total_MASsr)

```

a0	n	m0	total	total_est	desv_total_est	IC_inferior	IC_superior	coef_var_est
0.05	69	7811	2365.175	2280.226	107.4755	2069.578	2490.874	0.0471337
0.10	50	5348	2365.175	2149.897	215.5223	1727.481	2572.313	0.1002477
0.15	35	3541	2365.175	2012.608	325.1680	1375.291	2649.926	0.1615655

## 7 Técnica de media de medias con el MASsr para conglomerados

### 7.1 Cuasivarianza poblacional de y promedio

```

cuasivarianza_MediaMed <- function(data, n = nrow(data)) {
  promedio <- mean(data_conglomerado$Yi_prom)
  sumatoria <- c(NULL)
  for (i in 1:n) {
    sumatoria[i] <- (data$Yi_prom[i] -
                       promedio)^2
  }
  cuasivarianza <- sum(sumatoria) / (n - 1)
  cuasivarianza
}

```

### 7.2 Total estimado

$$\hat{Y}_1 = M_0 \frac{S\bar{y}_i}{n}$$

```

total_est_MediaMed <- function(data) {

  total_est <- mean(data$Yi_prom) * sum(data_conglomerado$Mi)
  total_est
}

```

### 7.3 Media estimada por elementos

$$\hat{Y}_{e1} = \frac{\hat{Y}_1}{M_0}$$

```
med_mues_ele_MediaMed <- function(data) {
  media_muestral_por_elemento <- total_est_MediaMed(data) / sum(data_conglomerado$Mi)
}
```

### 7.4 Estimada de la varianza de la media por elemento estimado

$$V(\hat{Y}_{e1}) = \left( \frac{1}{n} - \frac{1}{N} \right) s_{\bar{y}}^2$$

```
est_var_med_mues_el_MediaMed <- function(data, n = nrow(data), N = nrow(data_conglomerado)) {
  est_var <- ((1 / n) - (1 / N)) * cuasivarianza_MediaMed(data)
  est_var
}
```

### 7.5 Estimada de la varianza del total estimado

$$V(\hat{Y}_1) = M_0^2 * V(\hat{Y}_{e1})$$

```
est_var_total_est_MediaMed <- function(data) {
  est_var <- (sum(data_conglomerado$Mi)^2) *
    est_var_med_mues_el_MediaMed(data, n = nrow(data))
  est_var
}
```

### 7.6 Análisis del sesgo del estimador

```
sesgo_TotalEst_MediaMed <- function() {
  M_raya <- sum(data_conglomerado$Mi) / nrow(data_conglomerado)
  sumatoria <- c(NULL)
  for (i in 1:nrow(data_conglomerado)) {
    sumatoria[i] <- data_conglomerado$Yi_prom[i] *
      (M_raya - data_conglomerado$Mi[i])
  }
  sesgo <- sum(sumatoria)
  sesgo
}
```

```
relacion_sesgo_desv <- function(n) {

  numerador <- sesgo_TotalEst_MediaMed()
  a <- sum(data_conglomerado$Mi)^2
  b <- ((1 / n) - (1 / nrow(data_conglomerado)))
  c <- cuasivarianza_MediaMed(data = data_conglomerado)
  denominador <- a * b * c
  relacion <- abs(numerador) / sqrt(denominador)
```

```

    relacion
}

```

## 7.7 tamano de muestra mediante el coeficiente de variacion

```

tamano_de_muestra_MediaMed <- function(coeficiente) {

  M_raya <- sum(data_conglomerado$Mi) / nrow(data_conglomerado)
  CuasiPob <- cuasivarianza_MediaMed(data = data_conglomerado)
  numerador <- nrow(data_conglomerado) * CuasiPob * (sum(data_conglomerado$Mi)^2)
  denominador_a <- nrow(data_conglomerado) *
    ((coeficiente * M_raya * sum(data_conglomerado$Yi_prom))^2)
  denominador_b <- (sum(data_conglomerado$Mi)^2) * CuasiPob
  n <- numerador / (denominador_a + denominador_b)

  tabla <- tibble(a0 = coeficiente,
                  n, n_MediaMed = ceiling(n),
                  sesgo_sobre_desv = relacion_sesgo_desv(n))
  tabla
}

tabla_tamano_de_muestra_MediaMed <- tamano_de_muestra_MediaMed(coeficiente = a0)
kable(tabla_tamano_de_muestra_MediaMed,
      align = "cccc",
      caption = "Tamaño de muestra MediaMed con respecto al coeficiente de variacion")

```

Table 11: Tamaño de muestra MediaMed con respecto al coeficiente de variacion

a0	n	n_MediaMed	sesgo_sobre_desv
0.01	66.677965	67	7.0134210
0.02	46.450498	47	3.5067105
0.03	30.851798	31	2.3378070
0.04	20.985636	21	1.7533553
0.05	14.871182	15	1.4026842
0.06	10.966049	11	1.1689035
0.07	8.368842	9	1.0019173
0.08	6.572676	7	0.8766776
0.09	5.286723	6	0.7792690
0.10	4.338111	5	0.7013421
0.11	3.620160	4	0.6375837
0.12	3.064658	4	0.5844518
0.13	2.626573	3	0.5394939
0.14	2.275303	3	0.5009586
0.15	1.989522	2	0.4675614
0.16	1.754021	2	0.4383388
0.17	1.557735	2	0.4125542
0.18	1.392465	2	0.3896345
0.19	1.252038	2	0.3691274
0.20	1.131736	2	0.3506711

La relacion sesgo-desviacion estandar es mayor a 0.1 por lo cual se aconseja cambiar de tecnica de muestreo.

```
ggplot(tabla_tamano_de_muestra_MediaMed, aes(x = a0, y = n_MediaMed)) +
  geom_label(aes(label = n_MediaMed), colour = "blue", fontface = "bold") +
  geom_line() +
  geom_hline(yintercept = 78, color = "red") +
  annotate(
    "text",
    x = 0.1,
    y = 78,
    label = "78 = Numero de conglomerados en la poblacion",
    vjust = 1, size = 3, color = "red"
  ) +
  labs(x = "a0 = coeficiente de variacion",
       y = "n = Tamano de la muestra") +
  theme_stata()
```

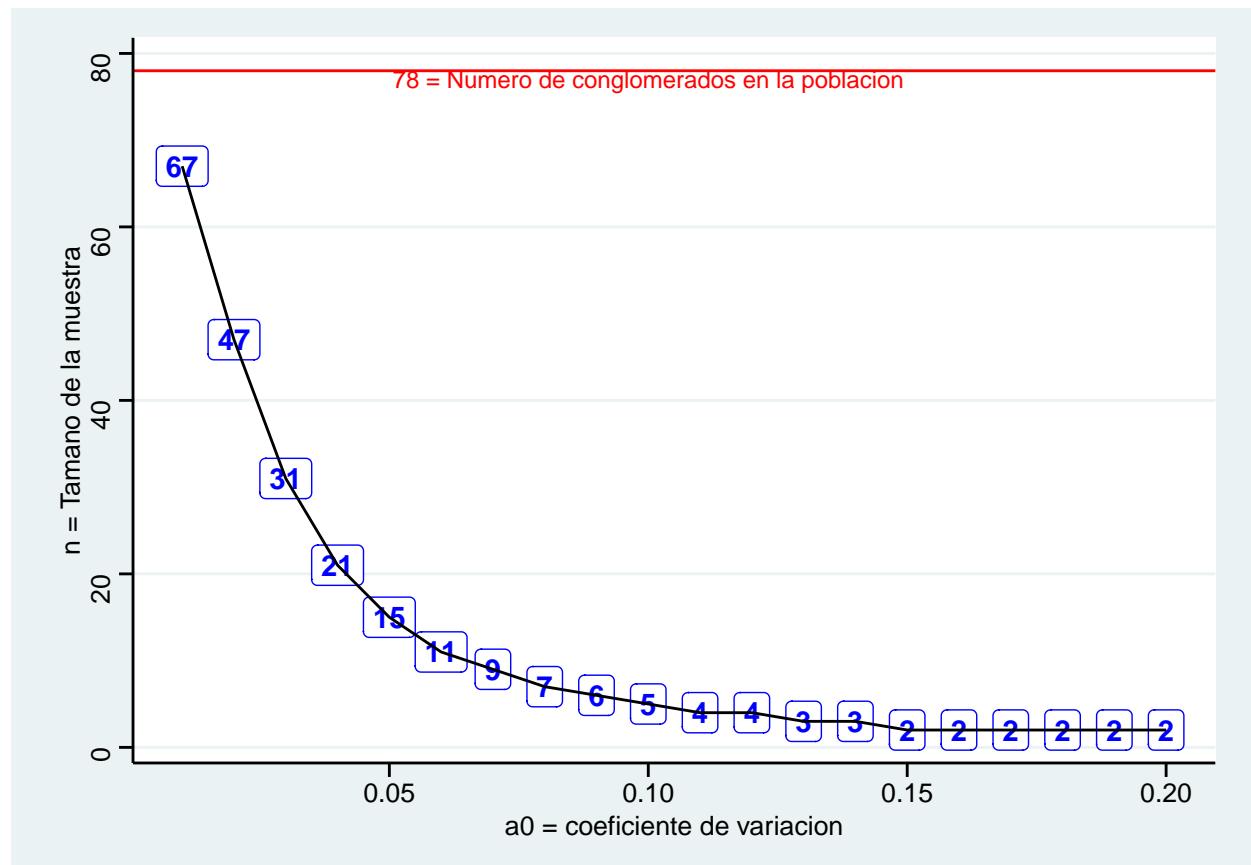


Figure 10: Grafico del tamano de muestra en Media de medias

## 7.8 Seleccion de las muestras

```
data_muestra_MediaMed <- function(n, N = nrow(data_conglomerado)) {

  set.seed(123)
  indice_MediaMed <- sample(1:N, size = n, replace = FALSE)
  data_MediaMed <- data_conglomerado[indice_MediaMed, 1:4]
```

```

data_MediaMed
}

```

## 7.9 Resultados Media de medias

```

data_15_MediaMed <- data_muestra_MediaMed(n = 15)
data_5_MediaMed <- data_muestra_MediaMed(n = 5)
data_2_MediaMed <- data_muestra_MediaMed(n = 2)

estimaciones_MediaMed <- function(data) {
  tibble(total_est = total_est_MediaMed(data),
         media_ele_est = med_mues_ele_MediaMed(data),
         var_media_est = est_var_med_mues_el_MediaMed(data),
         var_total_est = est_var_total_est_MediaMed(data))
}

estimaciones_data_15_MediaMed <- estimaciones_MediaMed(data = data_15_MediaMed)
estimaciones_data_5_MediaMed <- estimaciones_MediaMed(data = data_5_MediaMed)
estimaciones_data_2_MediaMed <- estimaciones_MediaMed(data = data_2_MediaMed)

```

## 7.10 Intervalo de confianza

$$(\hat{\bar{Y}}_{e1} - B \left( \hat{\bar{Y}}_{e1} \right) \pm t_{(1-\alpha/2;n-1)} \sqrt{V(\hat{\bar{Y}}_{e1})})$$

```

tabla_IC_media_MediaMed <-
  tibble(a0 = c(0.05, 0.10, 0.15),
         n = c(15, 5, 2),
         m0 = c(sum(data_15_MediaMed$Mi),
                sum(data_5_MediaMed$Mi),
                sum(data_2_MediaMed$Mi)),
         media_ele = c(mean(data_elemento$Yij),
                       mean(data_elemento$Yij),
                       mean(data_elemento$Yij)),
         media_ele_est = as.numeric(c(estimaciones_data_15_MediaMed[2],
                                       estimaciones_data_5_MediaMed[2],
                                       estimaciones_data_2_MediaMed[2])),
         desv_media_est = sqrt(as.numeric(c(estimaciones_data_15_MediaMed[3],
                                             estimaciones_data_5_MediaMed[3],
                                             estimaciones_data_2_MediaMed[3]))))

tabla_IC_media_MediaMed <-
  tabla_IC_media_MediaMed %>%
  mutate(IC_inferior = media_ele_est -
    (sesgo_TotalEst_MediaMed() / sum(data_conglomerado$Mi)) -
    (qt(p = 1-(0.05/2), df = n - 1) * desv_media_est),
    IC_superior = media_ele_est -
    (sesgo_TotalEst_MediaMed() / sum(data_conglomerado$Mi)) +
    (qt(p = 1-(0.05/2), df = n - 1) * desv_media_est),
    coef_var_est = desv_media_est / media_ele_est)

kable(tabla_IC_media_MediaMed)

```

a0	n	m0	media_ele	media_ele_est	desv_media_est	IC_inferior	IC_superior	coef_var_est
0.05	15	1094	0.2591973	0.2498730	0.0117610	0.2416353	0.2920851	0.0470680
0.10	5	251	0.2591973	0.2508983	0.0130475	0.2316600	0.3041110	0.0520029
0.15	2	128	0.2591973	0.2804541	0.0377972	-0.1828172	0.7776998	0.1347713

$$\langle \hat{Y}_1 - B(\hat{Y}_1) \pm t_{(1-\alpha/2;n-1)} \sqrt{\hat{V}(\hat{Y}_1)} \rangle$$

```

tabla_IC_total_MediaMed <-
  tibble(a0 = c(0.05, 0.10, 0.15),
  n = c(15, 5, 2),
  m0 = c(sum(data_15_MediaMed$Mi),
         sum(data_5_MediaMed$Mi),
         sum(data_2_MediaMed$Mi)),
  total = c(sum(data_conglomerado$Yi),
            sum(data_conglomerado$Yi),
            sum(data_conglomerado$Yi)),
  total_est = as.numeric(c(estimaciones_data_15_MediaMed[1],
                           estimaciones_data_5_MediaMed[1],
                           estimaciones_data_2_MediaMed[1])),
  desv_total_est = sqrt(as.numeric(c(estimaciones_data_15_MediaMed[4],
                                       estimaciones_data_5_MediaMed[4],
                                       estimaciones_data_2_MediaMed[4]))))

tabla_IC_total_MediaMed <-
  tabla_IC_total_MediaMed %>%
  mutate(IC_inferior = total_est - sesgo_TotalEst_MediaMed() -
    (qt(p = 1-(0.05/2), df = n - 1) * desv_total_est),
   IC_superior = total_est + sesgo_TotalEst_MediaMed() +
    (qt(p = 1-(0.05/2), df = n - 1) * desv_total_est),
   coef_var_est = desv_total_est / total_est)

kable(tabla_IC_total_MediaMed)

```

a0	n	m0	total	total_est	desv_total_est	IC_inferior	IC_superior	coef_var_est
0.05	15	1094	2365.175	2280.091	107.3194	2204.922	2665.276	0.0470680
0.10	5	251	2365.175	2289.447	119.0580	2113.897	2775.013	0.0520029
0.15	2	128	2365.175	2559.144	344.8991	-1668.207	7096.511	0.1347713

## 8 Técnica de razón para los conglomerados

### 8.1 Cuasivarianza de R estimado

$$S_{\hat{R}}^2 = \frac{\sum_{i=1}^n (y_i - \hat{R}M_i)^2}{n - 1}$$

```

cuasivarianza_Razon <- function(data, n = nrow(data)) {
  r <- mean(data$Yi) / mean(data$Mi)
  sumatoria <- c(NULL)
  for (i in 1:n) {
    sumatoria[i] <- (data$Yi[i] -
      (r * data$Mi[i]))^2
  }
  cuasivarianza <- sum(sumatoria) / (n - 1)
  cuasivarianza
}

varianza_Razon_est <- function(n, N = nrow(data_conglomerado)) {
  R <- sum(data_conglomerado$Yi) / sum(data_conglomerado$Mi)
  sumatoria <- c(NULL)
  for (i in 1:N) {
    sumatoria[i] <- (data_conglomerado$Yi[i] -
      (R * data_conglomerado$Mi[i]))^2
  }
  cuasivarianza <- sum(sumatoria) / (N - 1)
  a <- 1 - (n / N)
  b <- N * (mean(data_conglomerado$Mi))^2
  varianza <- a * cuasivarianza / b
  varianza
}

```

## 8.2 Total estimado

$$\hat{Y}_2 = X\hat{R} = M_0 \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n M_i}$$

```

total_est_Razon <- function(data) {

  total_est <- sum(data_conglomerado$Mi) * sum(data$Yi) / sum(data$Mi)
  total_est

}

```

## 8.3 Media estimada por elementos

$$\hat{Y}_{e2} = \frac{\hat{Y}_2}{M_0} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n M_i} = \hat{R}$$

```

med_mues_ele_Razon <- function(data) {
  media_muestral_por_elemento <- total_est_Razon(data) / sum(data_conglomerado$Mi)
}

```

## 8.4 Estimada de la varianza de la media por elemento estimado

$$V(\hat{Y}_{e2}) = \frac{N^2}{M_0^2} \frac{(1-f)}{n} \frac{\sum_{i=1}^n (y_i - \hat{R}M_i)^2}{n-1}$$

```

est_var_med_mues_el_Razon <- function(data, n = nrow(data), N = nrow(data_conglomerado)) {

```

```

a <- ((N^2) / (sum(data_conglomerado$Mi)^2))
b <- ((1 / n) - (1 / N)) * cuasivarianza_Razon(data)
est_var <- a * b
est_var

}

```

## 8.5 Estimada de la varianza del total estimado

$$V(\hat{Y}_2) = V(\hat{Y}_{e2}) * M_0^2$$

```

est_var_total_est_Razon <- function(data) {
  est_var <- (sum(data_conglomerado$Mi)^2) *
    est_var_med_mues_el_Razon(data)
  est_var
}

```

## 8.6 tamano de muestra mediante el coeficiente de variacion

$$n \geq \frac{NS_x^2}{a_0^2 * N\bar{X}^2 + S_x^2}$$

```

tamano_de_muestra_Razon <- function(coeficiente) {

  promedio <- mean(data_conglomerado$Mi)
  sumatoria <- c(NULL)
  for (i in 1:nrow(data_conglomerado)) {
    sumatoria[i] <- (data_conglomerado$Mi[i] -
      promedio)^2
  }

  CuasiPob_x <- sum(sumatoria) / (nrow(data_conglomerado) - 1)
  M_raya <- sum(data_conglomerado$Mi) / nrow(data_conglomerado)

  numerador <- nrow(data_conglomerado) * CuasiPob_x
  denominador <- (nrow(data_conglomerado) *
    ((coeficiente * M_raya)^2)) + CuasiPob_x
  n <- numerador / denominador

  n_Razon <- ceiling(n)

  desv_x <- sqrt((1 - (n_Razon / nrow(data_conglomerado))) * CuasiPob_x / n_Razon)
  cv_x <- desv_x / promedio

  tabla <- tibble(a0 = coeficiente,
                  n,
                  n_Razon,
                  cv_x = round(cv_x, digits = 3))
  tabla
}

}

```

```

tabla_tamano_de_muestra_Razon <- tamano_de_muestra_Razon(coeficiente = a0)
kable(tabla_tamano_de_muestra_Razon,
      align = "cccc",
      caption = "Tamaño de muestra Razon con respecto al coeficiente de variacion")

```

Table 14: Tamaño de muestra Razon con respecto al coeficiente de variacion

a0	n	n_Razon	cv_x
0.01	77.49098	78	0.000
0.02	76.00301	77	0.014
0.03	73.64611	74	0.029
0.04	70.58182	71	0.039
0.05	66.99767	67	0.050
0.06	63.08250	64	0.058
0.07	59.00730	60	0.068
0.08	54.91403	55	0.080
0.09	50.91147	51	0.090
0.10	47.07649	48	0.098
0.11	43.45834	44	0.108
0.12	40.08419	41	0.117
0.13	36.96465	37	0.130
0.14	34.09864	35	0.137
0.15	31.47730	32	0.148
0.16	29.08702	30	0.156
0.17	26.91161	27	0.170
0.18	24.93379	25	0.180
0.19	23.13627	24	0.185
0.20	21.50234	22	0.197

```

ggplot(tabla_tamano_de_muestra_Razon, aes(x = a0, y = cv_x)) +
  geom_label(aes(label = cv_x), colour = "blue", fontface = "bold") +
  geom_line() +
  geom_hline(yintercept = 0.1, color = "red") +
  annotate(
    "text",
    x = 0.15, y = 0.1,
    label = "0.1 = limite de aceptacion del sesgo",
    vjust = 1, size = 3, color = "red"
  ) +
  labs(x = "a0 = coeficiente de variacion",
       y = "cv(x) = CoefVar de la variable auxiliar") +
  theme_stata()

```

```

ggplot(tabla_tamano_de_muestra_Razon, aes(x = a0, y = n_Razon)) +
  geom_label(aes(label = n_Razon), colour = "blue", fontface = "bold") +
  geom_line() +
  geom_hline(yintercept = 78, color = "red") +
  annotate(
    "text",
    x = 0.1, y = 78,
    label = "78 = Numero de conglomerados en la poblacion",
  )

```

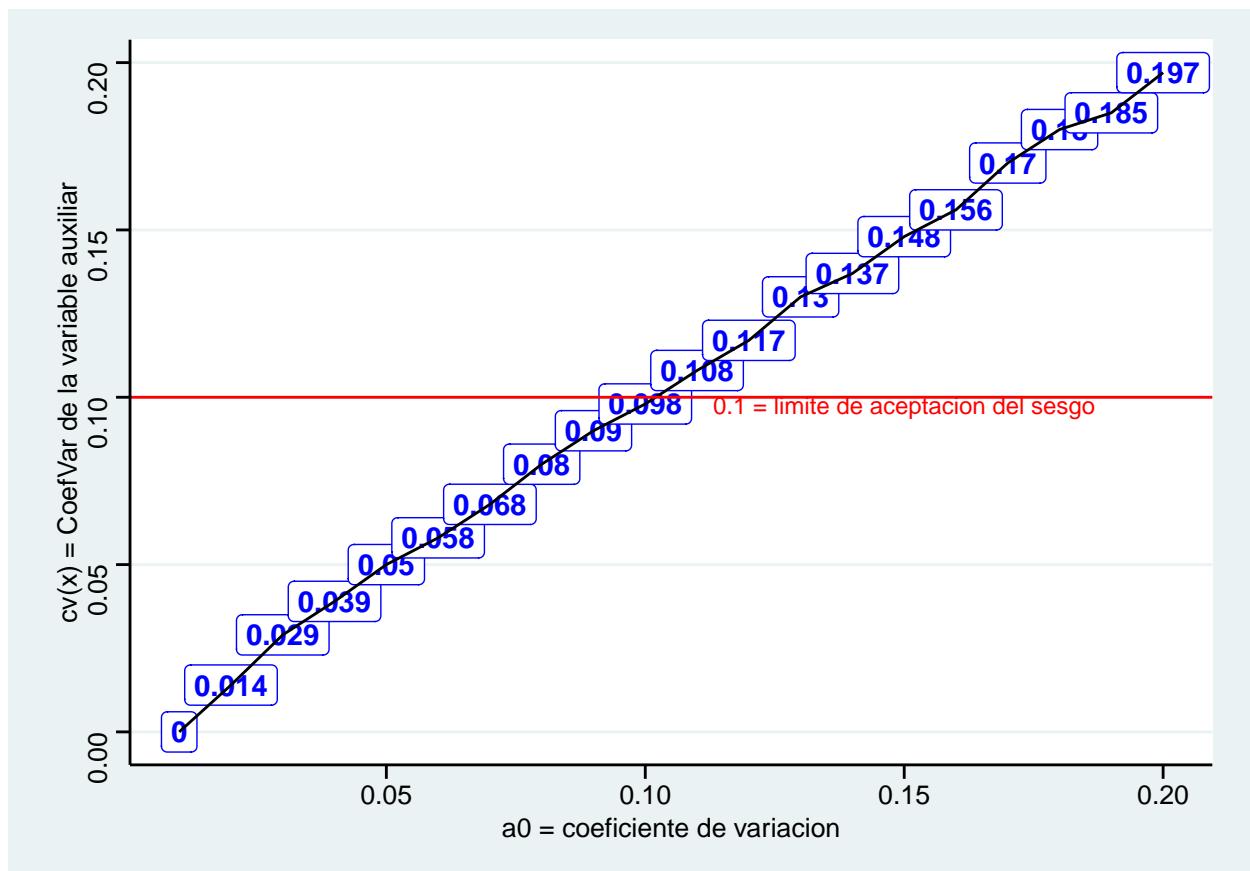


Figure 11: Grafico del  $cv(x)$  respecto al tamaño de muestra

```

    vjust = 1, size = 3, color = "red"
) +
labs(x = "a0 = coeficiente de variacion",
y = "n = Tamano de la muestra") +
theme_stata()

```

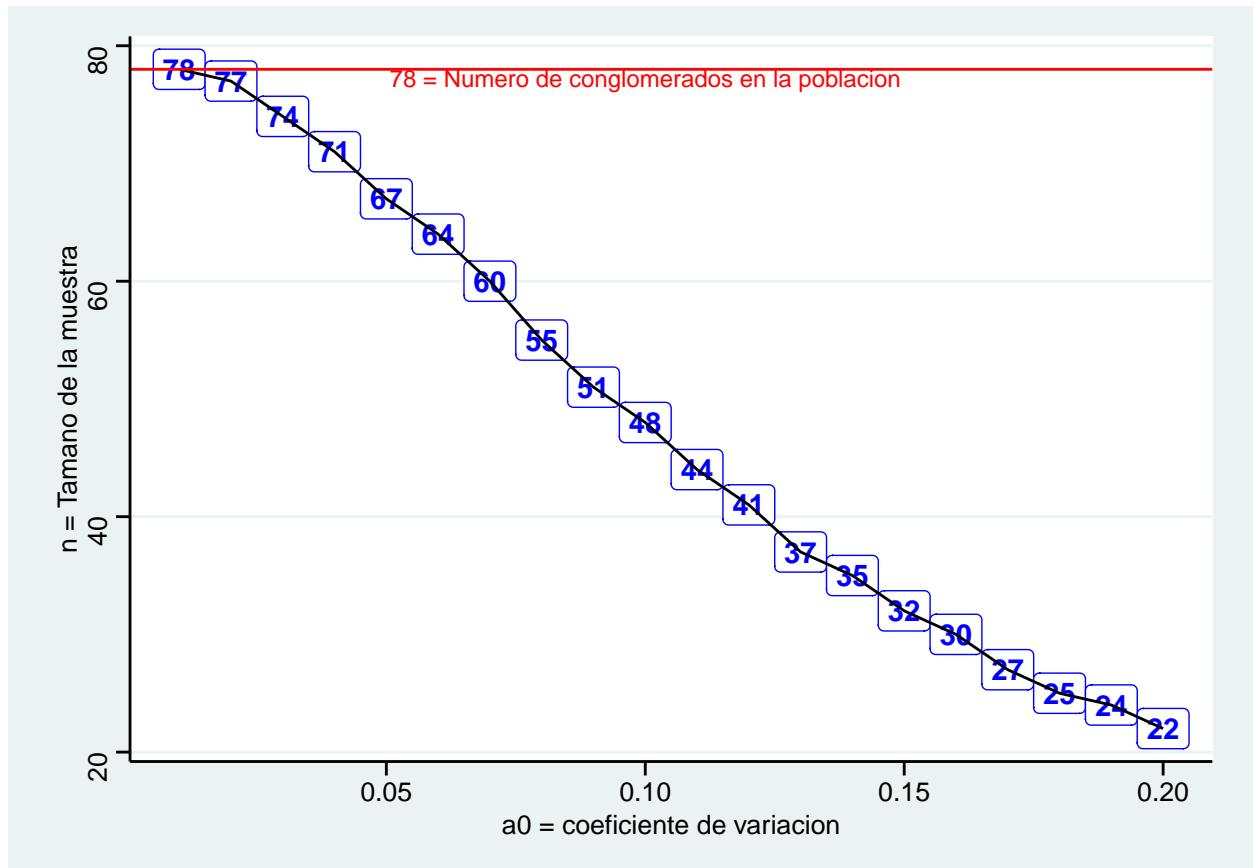


Figure 12: Grafico del tamano de muestra en Razon

## 8.7 Seleccion de las muestras

```

data_muestra_Razon <- function(n, N = nrow(data_conglomerado)) {

  set.seed(123)
  indice_Razon <- sample(1:N, size = n, replace = FALSE)
  data_Razon <- data_conglomerado[indice_Razon,1:3]

  data_Razon
}

```

## 8.8 Resultados Razon

```

data_67_Razon <- data_muestra_Razon(n = 67)
data_48_Razon <- data_muestra_Razon(n = 48)

```

```

data_32_Razon <- data_muestra_Razon(n = 32)

estimaciones_Razon <- function(data) {
  tibble(total_est = total_est_Razon(data),
         media_ele_est = med_mues_ele_Razon(data),
         var_media_est = est_var_med_mues_el_Razon(data),
         var_total_est = est_var_total_est_Razon(data))
}

estimaciones_data_67_Razon <- estimaciones_Razon(data = data_67_Razon)
estimaciones_data_48_Razon <- estimaciones_Razon(data = data_48_Razon)
estimaciones_data_32_Razon <- estimaciones_Razon(data = data_32_Razon)

```

## 8.9 Análisis del sesgo del estimador

$$\frac{|B(\hat{R})|}{\sigma_{\hat{R}}} = \frac{|-Cov(\hat{R}, \bar{x}) / \bar{X}|}{\sigma_{\hat{R}}}$$

```

sesgo_TotalEst_Razon <- function(data) {
  M_raya <- sum(data_conglomerado$Mi) / nrow(data_conglomerado)

  sesgo <- cov(data$Mi)
  sesgo
}

relacion_sesgo_desv <- function(n) {

  numerador <- sesgo_TotalEst_MediaMed()
  a <- sum(data_conglomerado$Mi)^2
  b <- ((1 / n) - (1 / nrow(data_conglomerado)))
  c <- cuasivarianza_MediaMed(data = data_conglomerado)
  denominador <- a * b * c
  relacion <- abs(numerador) / sqrt(denominador)
  relacion
}

```

## 8.10 Intervalo de confianza

```

tabla_IC_media_Razon <-
  tibble(a0 = c(0.05, 0.10, 0.15),
        n = c(67, 48, 32),
        m0 = c(sum(data_67_Razon$Mi),
               sum(data_48_Razon$Mi),
               sum(data_32_Razon$Mi)),
        media_ele = c(mean(data_elemento$Yij),
                     mean(data_elemento$Yij),
                     mean(data_elemento$Yij)),
        media_ele_est = as.numeric(c(estimaciones_data_67_Razon[2],
                                     estimaciones_data_48_Razon[2],
                                     estimaciones_data_32_Razon[2])),
        desv_media_est = sqrt(as.numeric(c(estimaciones_data_67_Razon[3],
                                         estimaciones_data_48_Razon[3],
                                         estimaciones_data_32_Razon[3])))

```

```

estimaciones_data_32_Razon[3]))))

tabla_IC_media_Razon <-
  tabla_IC_media_Razon %>%
  mutate(IC_inferior = media_ele_est -
    (qt(p = 1-(0.05/2), df = n - 1) * desv_media_est),
   IC_superior = media_ele_est +
    (qt(p = 1-(0.05/2), df = n - 1) * desv_media_est),
   coef_var_est = desv_media_est / media_ele_est,
   cociente = abs(-cov(media_ele_est, m0 / n) / mean(data_conglomerado$Mi)) /
    sqrt(varianza_Razon_est(n)))

kable(tabla_IC_media_Razon)

```

a0	n	m0	media_ele	media_ele	desv_media	est	IC_inferior	IC_superior	coef_var_est	cociente
0.05	67	7582	0.2591973	0.2590553	0.0024613	0.2541411	0.2639695	0.0095012	0.1350595	
0.10	48	5275	0.2591973	0.2588133	0.0048097	0.2491375	0.2684892	0.0185836	0.0817826	
0.15	32	3128	0.2591973	0.2514700	0.0080526	0.2350467	0.2678933	0.0320219	0.0660454	

```

tabla_IC_total_Razon <-
  tibble(a0 = c(0.05, 0.10, 0.15),
  n = c(67, 48, 32),
  m0 = c(sum(data_67_Razon$Mi),
         sum(data_48_Razon$Mi),
         sum(data_32_Razon$Mi)),
  total = c(sum(data_conglomerado$Yi),
            sum(data_conglomerado$Yi),
            sum(data_conglomerado$Yi)),
  total_est = as.numeric(c(estimaciones_data_67_Razon[1],
                           estimaciones_data_48_Razon[1],
                           estimaciones_data_32_Razon[1])),
  desv_total_est = sqrt(as.numeric(c(estimaciones_data_67_Razon[4],
                                      estimaciones_data_48_Razon[4],
                                      estimaciones_data_32_Razon[4]))))

tabla_IC_total_Razon <-
  tabla_IC_total_Razon %>%
  mutate(IC_inferior = total_est -
    (qt(p = 1-(0.05/2), df = n - 1) * desv_total_est),
   IC_superior = total_est +
    (qt(p = 1-(0.05/2), df = n - 1) * desv_total_est),
   coef_var_est = desv_total_est / total_est,
   cociente = abs(-cov(total_est, m0 / n) / mean(data_conglomerado$Mi)) /
    (sum(data_conglomerado$Mi) * sqrt(varianza_Razon_est(n)))))

kable(tabla_IC_total_Razon)

```

a0	n	m0	total	total_est	desv_total_est	IC_inferior	IC_superior	coef_var_est	cociente
0.05	67	7582	2365.175	2363.879	22.45964	2319.037	2408.722	0.0095012	0.1350595
0.10	48	5275	2365.175	2361.671	43.88846	2273.379	2449.964	0.0185836	0.0817826
0.15	32	3128	2365.175	2294.664	73.47957	2144.801	2444.526	0.0320219	0.0660454

## 9 Conclusiones

```

tabla_general <- tibble(a0 = c(c(0.05, 0.10, 0.15),
                               c(0.05, 0.10, 0.15),
                               c(0.05, 0.10, 0.15),
                               c(0.05, 0.10, 0.15)),
                           tecnica = c(c("MASsr", "MASsr", "MASsr"),
                                       c("MediaMedias", "MediaMedias", "MediaMedias"),
                                       c("PPTcr", "PPTcr", "PPTcr"),
                                       c("Razon", "Razon", "Razon")),
                           n = c(tabla_IC_media_MASsr$n,
                                 tabla_IC_media_MediaMed$n,
                                 tabla_IC_media_PPTcr$n,
                                 tabla_IC_media_Razon$n),
                           media_muestral = c(tabla_IC_media_MASsr$media_ele_est,
                                              tabla_IC_media_MediaMed$media_ele_est,
                                              tabla_IC_media_PPTcr$media_ele_est,
                                              tabla_IC_media_Razon$media_ele_est),
                           total_estimado = c(tabla_IC_total_MASsr$total_est,
                                              tabla_IC_total_MediaMed$total_est,
                                              tabla_IC_total_PPTcr$total_est,
                                              tabla_IC_total_Razon$total_est),
                           CoefVar_est = c(tabla_IC_total_MASsr$coef_var_est,
                                           tabla_IC_total_MediaMed$coef_var_est,
                                           tabla_IC_total_PPTcr$coef_var_est,
                                           tabla_IC_total_Razon$coef_var_est))

```

`kable(tabla_general)`

a0	tecnica	n	media_muestral	total_estimado	CoefVar_est
0.05	MASsr	69	0.2498878	2280.226	0.0471337
0.10	MASsr	50	0.2356052	2149.897	0.1002477
0.15	MASsr	35	0.2205598	2012.608	0.1615655
0.05	MediaMedias	15	0.2498730	2280.091	0.0470680
0.10	MediaMedias	5	0.2508983	2289.447	0.0520029
0.15	MediaMedias	2	0.2804541	2559.144	0.1347713
0.05	PPTcr	10	0.2503129	2284.105	0.0469576
0.10	PPTcr	3	0.2316928	2114.196	0.0923982
0.15	PPTcr	2	0.2470598	2254.421	0.1011360
0.05	Razon	67	0.2590553	2363.879	0.0095012
0.10	Razon	48	0.2588133	2361.671	0.0185836
0.15	Razon	32	0.2514700	2294.664	0.0320219

```

ggplot(tabla_general, aes(x = a0, y = n, color = tecnica)) +
  geom_text(aes(label = n), colour = "blue", fontface = "bold") +
  geom_line() +
  annotate(
    "text",
    x = 0.1, y = 78,
    label = "78 = Número de conglomerados en la población",
    vjust = 1, size = 3, color = "red"
  ) +

```

```

geom_hline(yintercept = 78, color = "red") +
labs(x = "a0 = coeficiente de variacion",
     y = "n = Tamano de la muestra") +
theme_stata()

```

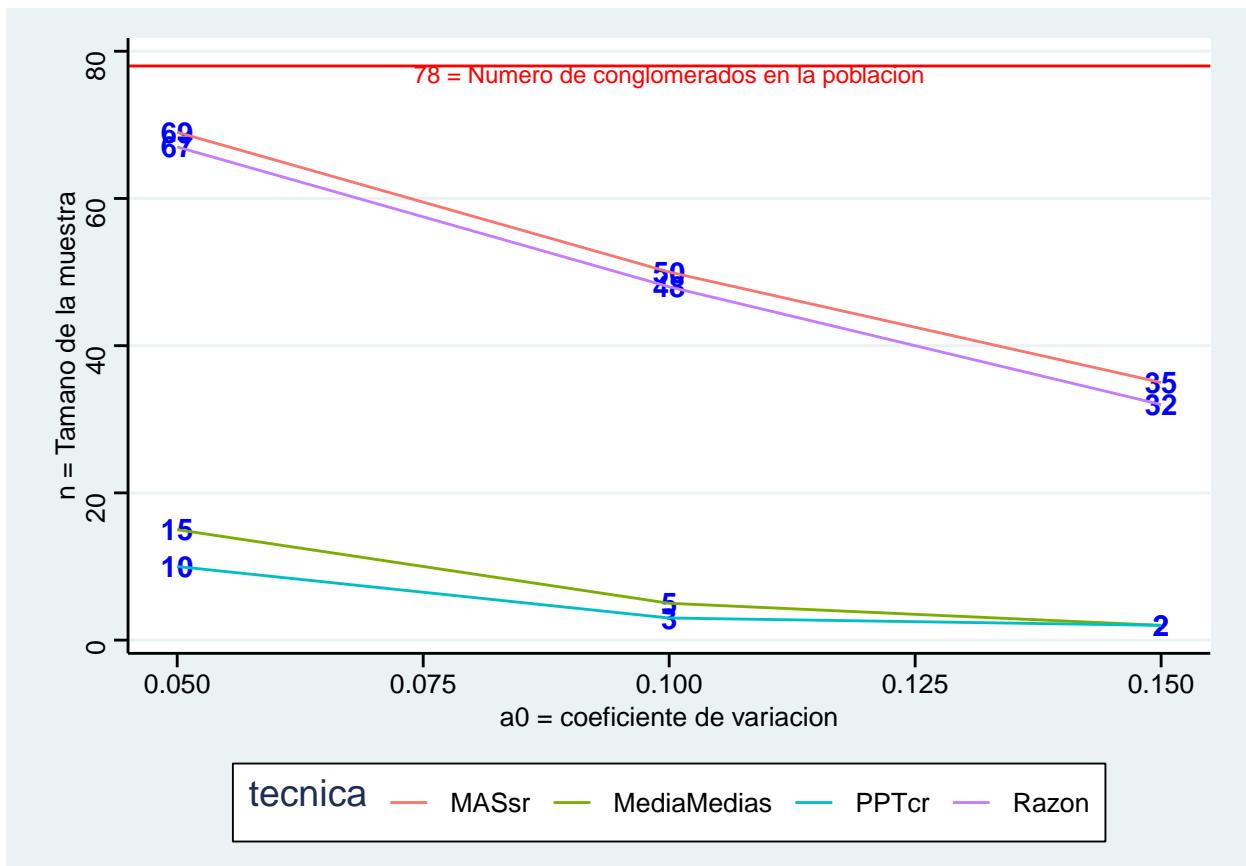


Figure 13: Grafico del tamano de muestras segun tecnicas

```

ggplot(tabla_general, aes(x = a0, y = round(CoefVar_est, digits = 2), color = tecnica)) +
  geom_text(aes(label = round(CoefVar_est, digits = 2)),
            colour = "blue", fontface = "bold") +
  geom_line() +
  annotate(
    "text",
    x = 0.075, y = 0.1,
    label = "0.1 = limite aceptable",
    vjust = 1, size = 5, color = "red"
  ) +
  geom_hline(yintercept = 0.1, color = "red") +
  labs(x = "a0 = coeficiente de variacion",
       y = "cv_est = coeficiente de variacion estimado") +
  theme_stata()

```

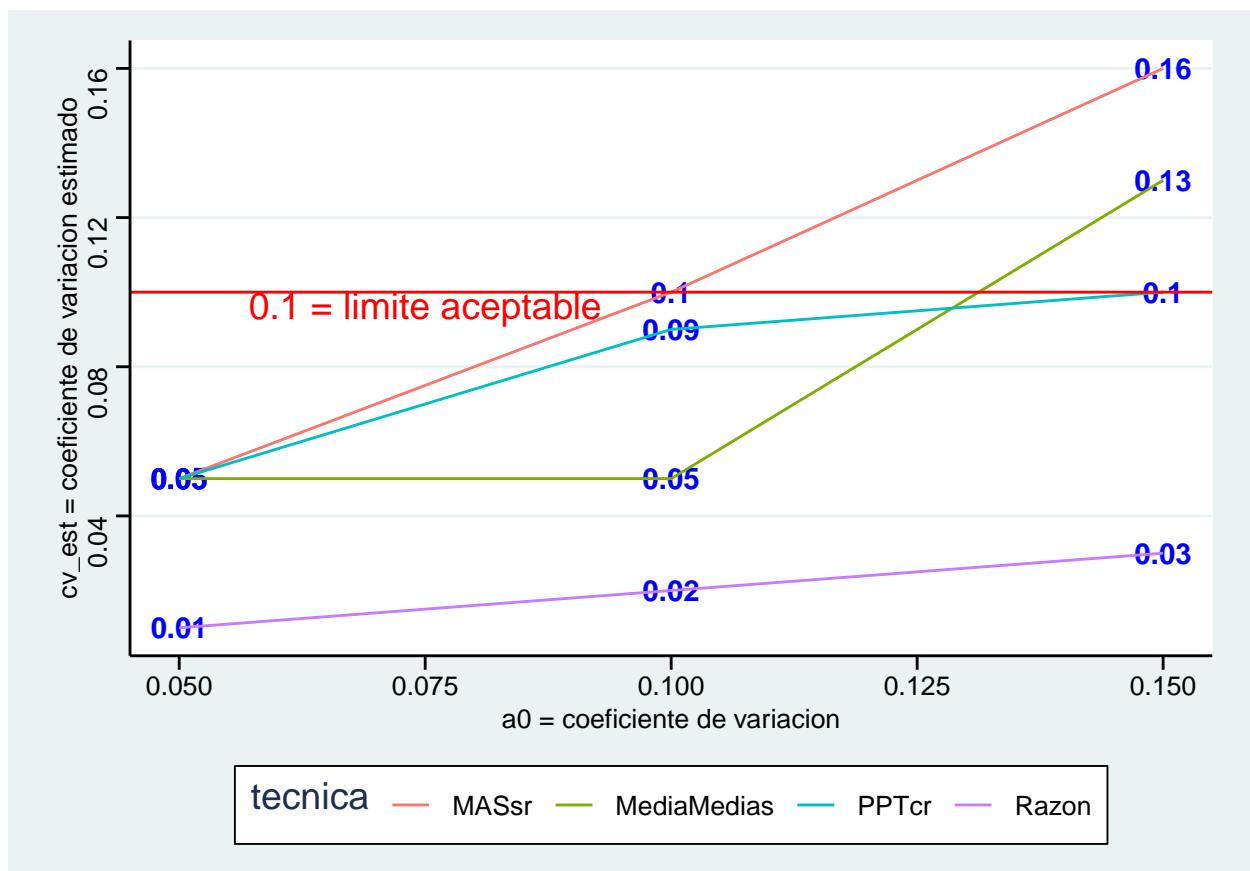


Figure 14: Grafico del coef\_var estimado segun tecnicas