

# Análisis de Encuestas Complejas (usando R)

## Lectura 4 - Estimación e Inferencia en Poblaciones Finitas: MAS - Estratificación - Conglomeración

Universidad de Santiago de Chile

Miguel Alvarado

November 23, 2020

# Outline

Muestreo Aleatorio Simple (MAS)

Muestreo Aleatorio Simple (MAS): Diseño complejo

Muestreo Aleatorio Simple (MAS): Estimación e Inferencia

Muestreo Estratificado: Diseño complejo - Estimación e Inferencia

Muestreo por Conglomerados: Diseño complejo - Estimación e Inferencia

# Muestreo Aleatorio Simple (MAS)

Un MAS de tamaño  $n$  de una población de tamaño  $N$ :

- $\pi_k = \frac{n}{N}, \forall k.$

# Muestreo Aleatorio Simple (MAS)

Un MAS de tamaño  $n$  de una población de tamaño  $N$ :

- $\pi_k = \frac{n}{N}, \forall k.$
- $d_k = \frac{1}{\pi_k} = \frac{N}{n}, \forall k.$

# Muestreo Aleatorio Simple (MAS)

Un MAS de tamaño  $n$  de una población de tamaño  $N$ :

- $\pi_k = \frac{n}{N}, \forall k.$
- $d_k = \frac{1}{\pi_k} = \frac{N}{n}, \forall k.$

Total de una variable  $y$ :  $t_y = \sum_{k \in U} y_k$

# Muestreo Aleatorio Simple (MAS)

Un MAS de tamaño  $n$  de una población de tamaño  $N$ :

- $\pi_k = \frac{n}{N}, \forall k.$
- $d_k = \frac{1}{\pi_k} = \frac{N}{n}, \forall k.$

Total de una variable  $y$ :  $t_y = \sum_{k \in U} y_k$

- El *estimador* de  $t_y$ :

$$\hat{t}_{y,\pi} = \sum_{k \in S} \frac{1}{\pi_k} y_k = \sum_{k \in S} d_k y_k = \frac{N}{n} \sum_{k \in S} y_k$$

# Muestreo Aleatorio Simple (MAS)

Un MAS de tamaño  $n$  de una población de tamaño  $N$ :

- $\pi_k = \frac{n}{N}, \forall k.$
- $d_k = \frac{1}{\pi_k} = \frac{N}{n}, \forall k.$

Total de una variable  $y$ :  $t_y = \sum_{k \in U} y_k$

- El *estimador* de  $t_y$ :

$$\hat{t}_{y,\pi} = \sum_{k \in S} \frac{1}{\pi_k} y_k = \sum_{k \in S} d_k y_k = \frac{N}{n} \sum_{k \in S} y_k$$

- La *varianza* de  $\hat{t}_{y,\pi}$ :

$$Var(\hat{t}_{y,\pi}) = \sum_{k,l \in S} \left( \frac{y_k y_l}{\pi_{kl}} - \frac{y_k}{\pi_k} \frac{y_l}{\pi_l} \right) = \frac{N-n}{N} \times N^2 \times \frac{Var(y)}{n}$$

# Muestreo Aleatorio Simple (MAS)

Un MAS de tamaño  $n$  de una población de tamaño  $N$ :

- $\pi_k = \frac{n}{N}, \forall k.$
- $d_k = \frac{1}{\pi_k} = \frac{N}{n}, \forall k.$

Total de una variable  $y$ :  $t_y = \sum_{k \in U} y_k$

- El *estimador* de  $t_y$ :

$$\hat{t}_{y,\pi} = \sum_{k \in S} \frac{1}{\pi_k} y_k = \sum_{k \in S} d_k y_k = \frac{N}{n} \sum_{k \in S} y_k$$

- La *varianza* de  $\hat{t}_{y,\pi}$ :

$$Var(\hat{t}_{y,\pi}) = \sum_{k,l \in S} \left( \frac{y_k y_l}{\pi_{kl}} - \frac{y_k}{\pi_k} \frac{y_l}{\pi_l} \right) = \frac{N-n}{N} \times N^2 \times \frac{Var(y)}{n}$$

- El *error estandar* de  $\hat{t}_{y,\pi}$ :

$$se(\hat{t}_{y,\pi}) = \sqrt{Var(\hat{t}_{y,\pi})}$$



# Muestreo Aleatorio Simple (MAS)

Corrección por Población Finita ( $fpc$ ):  $\frac{N-n}{N}$ , toma en cuenta la reducción en la incertidumbre (variabilidad) cuando una fracción grande de la población forma parte de la muestra.

# Muestreo Aleatorio Simple (MAS)

Corrección por Población Finita ( $fpc$ ):  $\frac{N-n}{N}$ , toma en cuenta la reducción en la incertidumbre (variabilidad) cuando una fracción grande de la población forma parte de la muestra.

- Imaginemos muestras independientes de tamaño  $n$  desde la misma población  $U$ :  $n/N$  es la proporción esperada que se sobrepone entre dos muestras cualesquiera y  $(N - n)/N$  es la fracción de la segunda muestra que previamente no fue observada. Entonces, la  $fpc$  representa la reducción en la variabilidad entre muestras, debido a que diferentes muestras incluye a las mismas observaciones.

# Muestreo Aleatorio Simple (MAS)

Corrección por Población Finita (*fpc*):  $\frac{N-n}{N}$ , toma en cuenta la reducción en la incertidumbre (variabilidad) cuando una fracción grande de la población forma parte de la muestra.

- Imaginemos muestras independientes de tamaño  $n$  desde la misma población  $U$ :  $n/N$  es la proporción esperada que se sobrepone entre dos muestras cualesquiera y  $(N - n)/N$  es la fracción de la segunda muestra que previamente no fue observada. Entonces, la *fpc* representa la reducción en la variabilidad entre muestras, debido a que diferentes muestras incluye a las mismas observaciones.
- No existe incertidumbre sobre los  $n$  individuos de la muestra, por tanto, solo es necesario estimar la incertidumbre en el total para los  $(N - n)$  individuos de la población que no están en la muestra.

# Muestreo Aleatorio Simple (MAS)

Corrección por Población Finita ( $fpc$ ):  $\frac{N-n}{N}$ , toma en cuenta la reducción en la incertidumbre (variabilidad) cuando una fracción grande de la población forma parte de la muestra.

- Imaginemos muestras independientes de tamaño  $n$  desde la misma población  $U$ :  $n/N$  es la proporción esperada que se sobrepone entre dos muestras cualesquiera y  $(N - n)/N$  es la fracción de la segunda muestra que previamente no fue observada. Entonces, la  $fpc$  representa la reducción en la variabilidad entre muestras, debido a que diferentes muestras incluye a las mismas observaciones.
- No existe incertidumbre sobre los  $n$  individuos de la muestra, por tanto, solo es necesario estimar la incertidumbre en el total para los  $(N - n)$  individuos de la población que no están en la muestra.
- Cuando  $n$  es muy pequeña respecto de  $N$ , la  $fpc$  suele ser ignorada.

# Muestreo Aleatorio Simple (MAS)

Corrección por Población Finita ( $fpc$ ):  $\frac{N-n}{N}$ , toma en cuenta la reducción en la incertidumbre (variabilidad) cuando una fracción grande de la población forma parte de la muestra.

- Imaginemos muestras independientes de tamaño  $n$  desde la misma población  $U$ :  $n/N$  es la proporción esperada que se sobrepone entre dos muestras cualesquiera y  $(N - n)/N$  es la fracción de la segunda muestra que previamente no fue observada. Entonces, la  $fpc$  representa la reducción en la variabilidad entre muestras, debido a que diferentes muestras incluye a las mismas observaciones.
- No existe incertidumbre sobre los  $n$  individuos de la muestra, por tanto, solo es necesario estimar la incertidumbre en el total para los  $(N - n)$  individuos de la población que no están en la muestra.
- Cuando  $n$  es muy pequeña respecto de  $N$ , la  $fpc$  suele ser ignorada.
- La  $fpc$  también desaparece de la fórmula si la muestra es tomada con reemplazo.

# Muestreo Aleatorio Simple (MAS)

Media de  $y$ :  $\mu_y = \frac{t_y}{N}$

# Muestreo Aleatorio Simple (MAS)

Media de  $y$ :  $\mu_y = \frac{t_y}{N}$

- El *estimador* de  $\mu_y$ :

$$\hat{\mu}_{y,\pi} = \frac{1}{N} \sum_{k \in S} d_k y_k = \frac{1}{n} \sum_{k \in S} y_k$$

# Muestreo Aleatorio Simple (MAS)

Media de  $y$ :  $\mu_y = \frac{t_y}{N}$

- El *estimador* de  $\mu_y$ :

$$\hat{\mu}_{y,\pi} = \frac{1}{N} \sum_{k \in S} d_k y_k = \frac{1}{n} \sum_{k \in S} y_k$$

- La *varianza* de  $\hat{\mu}_{y,\pi}$ :

$$Var(\hat{\mu}_{y,\pi}) = \frac{N - n}{N} \times \frac{Var(y)}{n}$$



# Muestreo Aleatorio Simple (MAS)

Media de  $y$ :  $\mu_y = \frac{t_y}{N}$

- El *estimador* de  $\mu_y$ :

$$\hat{\mu}_{y,\pi} = \frac{1}{N} \sum_{k \in S} d_k y_k = \frac{1}{n} \sum_{k \in S} y_k$$

- La *varianza* de  $\hat{\mu}_{y,\pi}$ :

$$Var(\hat{\mu}_{y,\pi}) = \frac{N - n}{N} \times \frac{Var(y)}{n}$$

El total de la población:  $N$

# Muestreo Aleatorio Simple (MAS)

Media de  $y$ :  $\mu_y = \frac{t_y}{N}$

- El *estimador* de  $\mu_y$ :

$$\hat{\mu}_{y,\pi} = \frac{1}{N} \sum_{k \in S} d_k y_k = \frac{1}{n} \sum_{k \in S} y_k$$

- La *varianza* de  $\hat{\mu}_{y,\pi}$ :

$$Var(\hat{\mu}_{y,\pi}) = \frac{N-n}{N} \times \frac{Var(y)}{n}$$

El total de la población:  $N$

- El *estimador* de  $N$ :

$$\hat{N}_\pi = \sum_{k \in S} \frac{1}{\pi_k}$$

# Muestreo Aleatorio Simple (MAS)

Media de  $y$ :  $\mu_y = \frac{t_y}{N}$

- El *estimador* de  $\mu_y$ :

$$\hat{\mu}_{y,\pi} = \frac{1}{N} \sum_{k \in S} d_k y_k = \frac{1}{n} \sum_{k \in S} y_k$$

- La *varianza* de  $\hat{\mu}_{y,\pi}$ :

$$Var(\hat{\mu}_{y,\pi}) = \frac{N-n}{N} \times \frac{Var(y)}{n}$$

El total de la población:  $N$

- El *estimador* de  $N$ :

$$\hat{N}_\pi = \sum_{k \in S} \frac{1}{\pi_k}$$

- El estimador pasado en  $\hat{N}_\pi$ , denominado *estimador ratio de la media* fué muy popular, pues  $N$  no siempre es conocido.

# Muestreo Aleatorio Simple (MAS)

Media de  $y$ :  $\mu_y = \frac{t_y}{N}$

- El *estimador* de  $\mu_y$ :

$$\hat{\mu}_{y,\pi} = \frac{1}{N} \sum_{k \in S} d_k y_k = \frac{1}{n} \sum_{k \in S} y_k$$

- La *varianza* de  $\hat{\mu}_{y,\pi}$ :

$$Var(\hat{\mu}_{y,\pi}) = \frac{N - n}{N} \times \frac{Var(y)}{n}$$

El total de la población:  $N$

- El *estimador* de  $N$ :

$$\hat{N}_\pi = \sum_{k \in S} \frac{1}{\pi_k}$$

- El estimador pasado en  $\hat{N}_\pi$ , denominado *estimador ratio de la media* fué muy popular, pues  $N$  no siempre es conocido.
- En encuestas de gran escala  $N$  es conocido y los  $d_k$  suelen ser ajustados de modo que  $\hat{N}_\pi = N$ .

# Muestreo Aleatorio Simple (MAS): Diseño complejo

```
# cargamos el paquete 'survey'  
library(survey)  
# cargamos las bases de datos a utilizar  
data(api)  
# base de datos: apisrs  
# help("apisrs")  
# describimos el diseño muestral MAS con fpc: 'srs_design'  
srs_design <- svydesign(id = ~1, fpc = ~fpc, data = apisrs)
```

# Muestreo Aleatorio Simple (MAS): Diseño complejo

```
# cargamos el paquete 'survey'
library(survey)
# cargamos las bases de datos a utilizar
data(api)
# base de datos: apisrs
# help("apisrs")
# describimos el diseño muestral MAS con fpc: 'srs_design'
srs_design <- svydesign(id = ~1, fpc = ~fpc, data = apisrs)
```

- 'svydesign()' función que toma la descripción del diseño muestral, agrega este a la base de datos y almacena ambas en un *objeto de diseño complejo*: 'srs\_design'.

# Muestreo Aleatorio Simple (MAS): Diseño complejo

```
# cargamos el paquete 'survey'
library(survey)
# cargamos las bases de datos a utilizar
data(api)
# base de datos: apisrs
# help("apisrs")
# describimos el diseño muestral MAS con fpc: 'srs_design'
srs_design <- svydesign(id = ~1, fpc = ~fpc, data = apisrs)
```

- 'svydesign()' función que toma la descripción del diseño muestral, agrega este a la base de datos y almacena ambas en un *objeto de diseño complejo*: 'srs\_design'.
- Al unir la información del diseño muestral y la base de datos en un solo objeto, que será con lo que trabajaremos en adelante, asegura que se usen de manera adecuada cuando estos se usan en conjunto y/o que la base de datos y un diseño muestral que pueda corresponder a otra base de datos, se usen en conjunto.

# Muestreo Aleatorio Simple (MAS): Diseño complejo

```
# Diseño MAS con FPC ("se conoce N")
```

```
srs_design
```

```
## Independent Sampling design
```

```
## svydesign(id = ~1, fpc = ~fpc, data = apisrs)
```



# Muestreo Aleatorio Simple (MAS): Diseño complejo

```
# Diseño MAS con FPC ("se conoce N")
```

```
srs_design
```

```
## Independent Sampling design
```

```
## svydesign(id = ~1, fpc = ~fpc, data = apisrs)
```

- 'id = ~ 1': dice que la selección es realizada sobre elementos individuales.
- 'fpc = ~ fpc': dice que la variable 'fpc' contiene el tamaño de la población.
- 'data = apisrs': dice que este diseño complejo le corresponde a la base de datos 'apisrs' y la información que ahí está contenida.

# Muestreo Aleatorio Simple (MAS): Diseño complejo

```
# Diseño MAS con FPC ("se conoce N")
```

```
srs_design
```

```
## Independent Sampling design
```

```
## svydesign(id = ~1, fpc = ~fpc, data = apisrs)
```

- 'id = ~ 1': dice que la selección es realizada sobre elementos individuales.
- 'fpc = ~ fpc': dice que la variable 'fpc' contiene el tamaño de la población.
- 'data = apisrs': dice que este diseño complejo le corresponde a la base de datos 'apisrs' y la información que ahí está contenida.
- Cuando se declara en 'fpc' el tamaño de la población, no es necesario incorporar los pesos de diseño (factores de expansión).

# Muestreo Aleatorio Simple (MAS): Diseño complejo

```
# Diseño MAS con FPC ("se conoce N")
```

```
srs_design
```

```
## Independent Sampling design
```

```
## svydesign(id = ~1, fpc = ~fpc, data = apisrs)
```

- 'id = ~ 1': dice que la selección es realizada sobre elementos individuales.
- 'fpc = ~ fpc': dice que la variable 'fpc' contiene el tamaño de la población.
- 'data = apisrs': dice que este diseño complejo le corresponde a la base de datos 'apisrs' y la información que ahí está contenida.
- Cuando se declara en 'fpc' el tamaño de la población, no es necesario incorporar los pesos de diseño (factores de expansión).
- Cuando el tamaño de la población no es especificado (en 'fpc'), es necesario especificar los pesos de diseño (factores de expansión) en 'weights = ~ pw'.

```
# describimos el diseño muestral MAS sin fpc (con pesos)
```

```
srs_design_noFPC <- svydesign(id = ~1, weights = ~pw, data = apisrs)
```

```
# Diseño MAS sin 'fpc' ("con reemplazo") y 'weights'
```

```
srs_design_noFPC
```

```
FALSE Independent Sampling design (with replacement)
```

```
FALSE svydesign(id = ~1, weights = ~pw, data = apisrs)
```

# Muestreo Aleatorio Simple (MAS): Estimación e Inferencia

Función 'svytotal()' estima el total poblacional y la precisión de la estimación.

```
# svytotal()
svytotal(~enroll, srs_design)

##           total      SE
## enroll 3621074 169520

svytotal(~enroll, srs_design_noFPC)

##           total      SE
## enroll 3621074 172325
```

Función 'svymean()' estima la media poblacional y la precisión de la estimación.

```
# svymean()
svymean(~enroll, srs_design, na.rm = TRUE)

##           mean      SE
## enroll 584.61 27.368

svymean(~enroll, srs_design_noFPC, na.rm = TRUE)

##           mean      SE
## enroll 584.61 27.821
```

# Muestreo Aleatorio Simple (MAS): Estimación e Inferencia

Las función 'svytotal()' y 'svymean()' se pueden aplicar en variables categóricas (factores): crea una tabla de de conteos y proporciones (respecto del total) estimadas para cada categoría de la variable.

```
# con variables categóricas  
svytotal(~stype, srs_design)
```

```
##           total      SE  
## stypeE 4397.74 196.00  
## stypeH  774.25 142.85  
## stypeM 1022.01 160.33
```

```
svymean(~stype, srs_design_noFPC)
```

```
##           mean      SE  
## stypeE 0.710 0.0322  
## stypeH 0.125 0.0234  
## stypeM 0.165 0.0263
```

# Muestreo Aleatorio Simple (MAS): Estimación e Inferencia

Utilizando 'svytotal()' y 'svymean()', múltiples variables pueden ser analizadas y realizar pruebas de contraste de diferencias.

```
# diferencias de medias
means <- svymean(~api00 + api99, srs_design)
means

##           mean      SE
## api00 656.59 9.2497
## api99 624.68 9.5003

svycontrast(means, c(api00 = 1, api99 = -1))

##           contrast      SE
## contrast      31.9 2.0905
```

Alternativamente:

```
# cuando creamos nuevas variables, actualizamos nuestro objeto de diseño complejo
srs_design <- update(srs_design, apidiff = api00 - api99)
srs_design <- update(srs_design, apipct = apidiff/api99)
svymean(~ apidiff + apipct, srs_design)

##           mean      SE
## apidiff 31.900000 2.0905
## apipct  0.056087 0.0041
```

# Muestreo Estratificado: Diseño complejo - Estimación e Inferencia

El MAS suele ser poco utilizado, porque otros diseños pueden dar (al menos) la misma precisión a un menor costo.

# Muestreo Estratificado: Diseño complejo - Estimación e Inferencia

El MAS suele ser poco utilizado, porque otros diseños pueden dar (al menos) la misma precisión a un menor costo.

El muestreo estratificado o estratificar la muestra es una manera de mejorar la precisión de las estimaciones: dividir la población en *estratos* y dentro de estos se realiza MAS independiente dentro de cada estrato (calculando probabilidades de selección por separado para cada estrato).



# Muestreo Estratificado: Diseño complejo - Estimación e Inferencia

El MAS suele ser poco utilizado, porque otros diseños pueden dar (al menos) la misma precisión a un menor costo.

El muestreo estratificado o estratificar la muestra es una manera de mejorar la precisión de las estimaciones: dividir la población en *estratos* y dentro de estos se realiza MAS independiente dentro de cada estrato (calculando probabilidades de selección por separado para cada estrato).

La estratificación, al asegurar muestra (observaciones) en cada estrato, hace que la muestra sea menos variable y esto mejora la precisión de las estimaciones.

# Muestreo Estratificado: Diseño complejo - Estimación e Inferencia

El MAS suele ser poco utilizado, porque otros diseños pueden dar (al menos) la misma precisión a un menor costo.

El muestreo estratificado o estratificar la muestra es una manera de mejorar la precisión de las estimaciones: dividir la población en *estratos* y dentro de estos se realiza MAS independiente dentro de cada estrato (calculando probabilidades de selección por separado para cada estrato).

La estratificación, al asegurar muestra (observaciones) en cada estrato, hace que la muestra sea menos variable y esto mejora la precisión de las estimaciones.

```
# describimos el diseño muestral Estratificado (y MAS): 'strat_design'  
strat_design <- svydesign(id = ~1, strata = ~stype, fpc = ~fpc, data = apistrat)  
strat_design
```

```
FALSE Stratified Independent Sampling design  
FALSE svydesign(id = ~1, strata = ~stype, fpc = ~fpc, data = apistrat)
```

- 'strata = ~ stype': dice que la variable de estratificación es 'stype'.

# Muestreo Estratificado: Diseño complejo - Estimación e Inferencia

Comparemos los resultados para un mismo estudio, pero con diseños distintos: MAS vs Estratificado.

```
# Estratificado vs MAS: total  
svytotal(~enroll, srs_design)
```

```
FALSE          total      SE  
FALSE enroll 3621074 169520
```

```
svytotal(~enroll, strat_design)
```

```
FALSE          total      SE  
FALSE enroll 3687178 114642
```

# Muestreo Estratificado: Diseño complejo - Estimación e Inferencia

Comparemos los resultados para un mismo estudio, pero con diseños distintos: MAS vs Estratificado.

```
# Estratificado vs MAS: total  
svytotal(~enroll, srs_design)
```

```
FALSE          total      SE  
FALSE enroll 3621074 169520
```

```
svytotal(~enroll, strat_design)
```

```
FALSE          total      SE  
FALSE enroll 3687178 114642
```

```
# Estratificado vs MAS: media  
svymean(~enroll, srs_design)
```

```
FALSE          mean      SE  
FALSE enroll 584.61 27.368
```

```
svymean(~enroll, strat_design)
```

```
FALSE          mean      SE  
FALSE enroll 595.28 18.509
```

# Muestreo Estratificado: Diseño complejo - Estimación e Inferencia

Para estratificar se debe tener información adicional sobre la población: tipo de escuela (stype)

```
# Estimación para la variable de estratificación  
svytotal(~stype, strat_design)
```

```
FALSE          total SE  
FALSE stypeE    4421  0  
FALSE stypeH     755  0  
FALSE stypeM    1018  0
```

# Muestreo Estratificado: Diseño complejo - Estimación e Inferencia

Función 'svyquantile()' estima cuantiles y la precisión de estos.

```
# svyquantile()
svyquantile(~api00, srs_design, c(0.25, 0.50, 0.75), ci = TRUE)

FALSE $quantiles
FALSE      0.25 0.5 0.75
FALSE api00  544 658 752
FALSE
FALSE $CIs
FALSE , , api00
FALSE
FALSE      0.25      0.5      0.75
FALSE (lower 528.0856 631.6647 738.1635
FALSE upper) 572.6576 692.3353 784.5482

svyquantile(~api00, strat_design, c(0.25, 0.50, 0.75), ci = TRUE)

FALSE $quantiles
FALSE      0.25      0.5      0.75
FALSE api00 562.2056 667.2358 755.1226
FALSE
FALSE $CIs
FALSE , , api00
FALSE
FALSE      0.25      0.5      0.75
FALSE (lower 534.0000 636.4618 725.1052
FALSE upper) 594.2887 681.0000 776.9814
```

# Muestreo por Conglomerados: Diseño complejo - Estimación e Inferencia

Un muestreo estratificado es factible en entrevistas telefónicas o cuestionarios por correo (los costos logísticos son bajos), pero cuando los cuestionarios deben aplicarse persona a persona, existe el problema que la muestra está repartida sobre la población entera, la que puede estar distante entre personas.

# Muestreo por Conglomerados: Diseño complejo - Estimación e Inferencia

Un muestreo estratificado es factible en entrevistas telefónicas o cuestionarios por correo (los costos logísticos son bajos), pero cuando los cuestionarios deben aplicarse persona a persona, existe el problema que la muestra está repartida sobre la población entera, la que puede estar distante entre personas.

El muestreo por conglomerados es una manera de reducir estos costos, para investigar un mismo tamaño de muestra.



# Muestreo por Conglomerados: Diseño complejo - Estimación e Inferencia

Un muestreo estratificado es factible en entrevistas telefónicas o cuestionarios por correo (los costos logísticos son bajos), pero cuando los cuestionarios deben aplicarse persona a persona, existe el problema que la muestra está repartida sobre la población entera, la que puede estar distante entre personas.

El muestreo por conglomerados es una manera de reducir estos costos, para investigar un mismo tamaño de muestra.

El muestreo por conglomerados reduce la precisión para un tamaño específico de muestra, pero puede incrementar el tamaño de la muestra y la precisión para un costo específico.

# Muestreo por Conglomerados: Diseño complejo - Estimación e Inferencia

El muestreo por conglomerados puede utilizarse en diseños de una sola etapa: toda la población del conglomerado seleccionado es investigada.

# Muestreo por Conglomerados: Diseño complejo - Estimación e Inferencia

El muestreo por conglomerados puede utilizarse en diseños de una sola etapa: toda la población del conglomerado seleccionado es investigada. Sin embargo, es más común en diseños de más de una etapa: Se seleccionan conglomerados de una primera etapa (PSU), de estos se seleccionan conglomerados de una segunda etapa (SSU), sucesivamente (las probabilidades de selección se calculan como el producto de las probabilidades de selección de cada etapa).

# Muestreo por Conglomerados: Diseño complejo - Estimación e Inferencia

El muestreo por conglomerados puede utilizarse en diseños de una sola etapa: toda la población del conglomerado seleccionado es investigada. Sin embargo, es más común en diseños de más de una etapa: Se seleccionan conglomerados de una primera etapa (PSU), de estos se seleccionan conglomerados de una segunda etapa (SSU), sucesivamente (las probabilidades de selección se calculan como el producto de las probabilidades de selección de cada etapa).

```
# describimos el diseño muestral por conglomerados en múltiples etapas: clus2_design  
clus2_design <- svydesign(id = ~dnum+snum, fpc = ~fpc1+fpc2, data = apiclus2)  
clus2_design
```

FALSE 2 - level Cluster Sampling design

FALSE With (40, 126) clusters.

```
FALSE svydesign(id = ~dnum + snum, fpc = ~fpc1 + fpc2, data = apiclus2)
```

# Muestreo por Conglomerados: Diseño complejo - Estimación e Inferencia

El muestreo por conglomerados puede utilizarse en diseños de una sola etapa: toda la población del conglomerado seleccionado es investigada. Sin embargo, es más común en diseños de más de una etapa: Se seleccionan conglomerados de una primera etapa (PSU), de estos se seleccionan conglomerados de una segunda etapa (SSU), sucesivamente (las probabilidades de selección se calculan como el producto de las probabilidades de selección de cada etapa).

```
# describimos el diseño muestral por conglomerados en múltiples etapas: clus2_design  
clus2_design <- svydesign(id = ~dnum+snum, fpc = ~fpc1+fpc2, data = apiclus2)  
clus2_design
```

FALSE 2 - level Cluster Sampling design

FALSE With (40, 126) clusters.

```
FALSE svydesign(id = ~dnum + snum, fpc = ~fpc1 + fpc2, data = apiclus2)
```

El muestreo por conglomerados en una o más etapas puede combinarse con el muestreo estratificado. Sin embargo, surgen problemas cuando existen estratos con solo un conglomerado.

# Muestreo por Conglomerados: Diseño complejo - Estimación e Inferencia

El muestreo por conglomerados puede utilizarse en diseños de una sola etapa: toda la población del conglomerado seleccionado es investigada. Sin embargo, es más común en diseños de más de una etapa: Se seleccionan conglomerados de una primera etapa (PSU), de estos se seleccionan conglomerados de una segunda etapa (SSU), sucesivamente (las probabilidades de selección se calculan como el producto de las probabilidades de selección de cada etapa).

```
# describimos el diseño muestral por conglomerados en múltiples etapas: clus2_design  
clus2_design <- svydesign(id = ~dnum+snum, fpc = ~fpc1+fpc2, data = apiclus2)  
clus2_design
```

FALSE 2 - level Cluster Sampling design

FALSE With (40, 126) clusters.

```
FALSE svydesign(id = ~dnum + snum, fpc = ~fpc1 + fpc2, data = apiclus2)
```

El muestreo por conglomerados en una o más etapas puede combinarse con el muestreo estratificado. Sin embargo, surgen problemas cuando existen estratos con solo un conglomerado.

- `options(survey.lonely.psu="adjust")`

# Muestreo por Conglomerados: Diseño complejo - Estimación e Inferencia

El muestreo por conglomerados puede utilizarse en diseños de una sola etapa: toda la población del conglomerado seleccionado es investigada. Sin embargo, es más común en diseños de más de una etapa: Se seleccionan conglomerados de una primera etapa (PSU), de estos se seleccionan conglomerados de una segunda etapa (SSU), sucesivamente (las probabilidades de selección se calculan como el producto de las probabilidades de selección de cada etapa).

```
# describimos el diseño muestral por conglomerados en múltiples etapas: clus2_design  
clus2_design <- svydesign(id = ~dnum+snum, fpc = ~fpc1+fpc2, data = apiclus2)  
clus2_design
```

FALSE 2 - level Cluster Sampling design

FALSE With (40, 126) clusters.

```
FALSE svydesign(id = ~dnum + snum, fpc = ~fpc1 + fpc2, data = apiclus2)
```

El muestreo por conglomerados en una o más etapas puede combinarse con el muestreo estratificado. Sin embargo, surgen problemas cuando existen estratos con solo un conglomerado.

- `options(survey.lonely.psu="adjust")`
- `options(survey.lonely.psu="average")`

# Muestreo por Conglomerados: Diseño complejo - Estimación e Inferencia

El muestreo por conglomerados puede utilizarse en diseños de una sola etapa: toda la población del conglomerado seleccionado es investigada. Sin embargo, es más común en diseños de más de una etapa: Se seleccionan conglomerados de una primera etapa (PSU), de estos se seleccionan conglomerados de una segunda etapa (SSU), sucesivamente (las probabilidades de selección se calculan como el producto de las probabilidades de selección de cada etapa).

```
# describimos el diseño muestral por conglomerados en múltiples etapas: clus2_design  
clus2_design <- svydesign(id = ~dnum+snum, fpc = ~fpc1+fpc2, data = apiclus2)  
clus2_design
```

FALSE 2 - level Cluster Sampling design

FALSE With (40, 126) clusters.

```
FALSE svydesign(id = ~dnum + snum, fpc = ~fpc1 + fpc2, data = apiclus2)
```

El muestreo por conglomerados en una o más etapas puede combinarse con el muestreo estratificado. Sin embargo, surgen problemas cuando existen estratos con solo un conglomerado.

- `options(survey.lonely.psu="adjust")`
- `options(survey.lonely.psu="average")`
- `options(survey.lonely.psu="remove")`



# Muestreo por Conglomerados: Diseño complejo - Estimación e Inferencia

El muestreo por conglomerados puede utilizarse en diseños de una sola etapa: toda la población del conglomerado seleccionado es investigada. Sin embargo, es más común en diseños de más de una etapa: Se seleccionan conglomerados de una primera etapa (PSU), de estos se seleccionan conglomerados de una segunda etapa (SSU), sucesivamente (las probabilidades de selección se calculan como el producto de las probabilidades de selección de cada etapa).

```
# describimos el diseño muestral por conglomerados en múltiples etapas: clus2_design  
clus2_design <- svydesign(id = ~dnum+snum, fpc = ~fpc1+fpc2, data = apiclus2)  
clus2_design
```

FALSE 2 - level Cluster Sampling design

FALSE With (40, 126) clusters.

```
FALSE svydesign(id = ~dnum + snum, fpc = ~fpc1 + fpc2, data = apiclus2)
```

El muestreo por conglomerados en una o más etapas puede combinarse con el muestreo estratificado. Sin embargo, surgen problemas cuando existen estratos con solo un conglomerado.

- `options(survey.lonely.psu="adjust")`
- `options(survey.lonely.psu="average")`
- `options(survey.lonely.psu="remove")`
- `options(survey.lonely.psu="certainty")`