

Análisis de encuestas de hogares con R

Agregación de encuestas

CEPAL - Unidad de Estadísticas Sociales

Tabla de contenidos I

Introducción

Factores de expansión y estimadores de muestreo

Agregación de encuestas con diferentes tamaños de muestra

Efecto del tipo de encuesta en la eficiencia de los indicadores

Pruebas de hipótesis sobre indicadores agregados

Introducción

¿Por qué agregar encuestas?

- ▶ Las encuestas de hogares suelen tener una frecuencia mensual o trimestral, pero los usuarios requieren *indicadores agregados* (anuales, semestrales).
- ▶ La agregación mejora la precisión estadística y permite el análisis de *subgrupos poblacionales*.
- ▶ Sin embargo, esta práctica plantea retos técnicos, como:
 - ▶ Duplicidad de unidades por esquemas rotativos.
 - ▶ Corrección de pesos de diseño muestral.
 - ▶ Estimación conjunta sin sesgos por repetición.

Fundamentos metodológicos

- ▶ Desde los años 60 se proponen enfoques basados en:
 - ▶ *Correlaciones temporales* entre periodos repetidos (Gurney y Daly (1965)).
 - ▶ *Estimadores compuestos* optimizados (Lent, Miller, y Duff (1999)).
- ▶ Se reconocen riesgos de:
 - ▶ Sesgos por error de medición (Fuller (1990)).
 - ▶ Desbalance muestral por rotación (Bell (2001)).

La agregación requiere ajustes tanto en los estimadores de nivel como en los estimadores de cambio.

Retos y buenas prácticas

- ▶ Considerar el efecto de la *rotación de paneles*:
 - ▶ Determina varianzas y errores estándar.
 - ▶ Afecta la independencia entre observaciones.
- ▶ Incorporar procedimientos para:
 - ▶ Combinar múltiples rondas sin inflar varianzas.
 - ▶ Ajustar pesos y corregir por no respuesta o traslape.
 - ▶ Aumentar potencia estadística en subgrupos (Lewis (2017)).
- ▶ Diseño robusto + estimación adecuada = *Indicadores válidos y eficientes*

¿Qué son los esquemas de acumulación?

- ▶ Utilizados por oficinas estadísticas para incrementar precisión y lograr representatividad a nivel subnacional.
- ▶ Consisten en agrupar levantamientos de encuestas (mensuales o trimestrales) para estimar indicadores agregados:
 - ▶ Ejemplo.: crear un indicador anual a partir de 4 levantamientos trimestrales.
- ▶ Estrategia útil especialmente en encuestas con diseño rotativo.

¿Qué son los esquemas de acumulación?

- ▶ Utilizados por oficinas estadísticas para incrementar precisión y lograr representatividad a nivel subnacional.
- ▶ Consisten en agrupar levantamientos de encuestas (mensuales o trimestrales) para estimar indicadores agregados:
 - ▶ Ejemplo.: crear un indicador anual a partir de 4 levantamientos trimestrales.
- ▶ Estrategia útil especialmente en encuestas con diseño rotativo.

Ideal para generar estimaciones provinciales o de subgrupos no representados en una sola ronda.

Trimestres móviles y acumulación anual

- ▶ Práctica común: usar trimestres móviles (últimos 3 meses) para publicar cifras mensuales de empleo.
- ▶ Ventajas en diseños rotativos:
 - ▶ Se mantiene el panel sin repetir vivienda en un mismo trimestre móvil.
 - ▶ En acumulación anual sí se repiten viviendas → requiere atención estadística.
- ▶ Ejemplo.: diseño 2(2)2 implica 50% de traslape entre muestras consecutivas.

La acumulación no es solo sumar bases: requiere estrategia de diseño y ponderación.

Consideraciones técnicas y desafíos

- ▶ Requiere codificación consistente de UPMs y estratos:
 1. En paneles → mismas UPM deben tener mismo código.
 2. En muestras independientes → codificar UPMs como distintas.
- ▶ Estimar varianzas es complejo:
 - ▶ Depende del patrón de traslape.
 - ▶ Solución práctica: métodos de remuestreo (*Jackknife*, *Bootstrap*, *BRR*).
- ▶ Referentes clave:
 - ▶ Korn y Graubard (1999) ponderación y diseño.
 - ▶ Train, Cahoon, y Makens (1978): variabilidad en promedios acumulados.

Factores de expansión y estimadores de muestreo

Factores de expansión y estimadores de muestreo

- ▶ Para estimar indicadores anuales, como la tasa de desempleo, en encuestas rotativas, es posible unir los datos de los cuatro trimestres del año.
- ▶ Una solución inicial consiste en agregar las cuatro bases de datos y dividir los pesos por 4.
- ▶ Esto produce estimadores puntuales aproximadamente insesgados.
- ▶ Sin embargo, complica la estimación de errores estándar debido a la concatenación de UPMs.

Tipos de pesos en encuestas rotativas

- ▶ Las encuestas rotativas requieren dos tipos de pesos:
 - ▶ **Transversales**: aplicables a cada periodo individual (mes, trimestre, etc.).
 - ▶ **Agregados**: ajustados para la combinación de varios periodos.
- ▶ Los pesos transversales se usan para:
 - ▶ Tasa de participación laboral.
 - ▶ Tasa de pobreza.
 - ▶ Tasa de desempleo.

Estimador de razón: Tasa de desempleo

- ▶ Se utiliza un estimador de razón:

$$\hat{\theta} = \frac{\sum_s d_k y_k}{\sum_s d_k z_k}$$

Donde:

- ▶ d_k : peso de muestreo para la persona k
- ▶ $y_k = 1$ si la persona está desempleada
- ▶ $z_k = 1$ si la persona pertenece a la fuerza laboral

Variabilidad de los pesos transversales

- ▶ Los pesos pueden cambiar entre periodos por:
 - ▶ Selección de UPM y hogares
 - ▶ Ajustes por no respuesta
 - ▶ Cambios en población objetivo
 - ▶ Calibración
- ▶ Por tanto, es común que:

$$d_k^{t-1} \neq d_k^t$$

- ▶ Esto exige cautela al construir pesos agregados para análisis acumulados.

Construcción de pesos agregados

- ▶ Es necesario crear nuevos factores de expansión para soportar la inferencia en bases agregadas.
- ▶ Cada peso en encuestas mensuales representa la cantidad de hogares que representa el hogar seleccionado.
- ▶ Para mantener la coherencia poblacional, se ajustan los pesos de forma proporcional.

Agregación trimestral

- ▶ Sea d_{ik} el factor de expansión para el individuo k en el mes i , con $i = 1, 2, 3$.
- ▶ La estimación total trimestral se aproxima por:

$$\hat{t}_y = \sum_{s_1 \cup s_2 \cup s_3} d_k^+ y_k \propto \sum_{s_1} d_{1k} y_k + \sum_{s_2} d_{2k} y_k + \sum_{s_3} d_{3k} y_k$$

- ▶ El peso agregado se obtiene mediante:

$$d_{ik}^+ = a_i \cdot d_{ik}; \quad k \in s_i$$

donde:

$$a_i = \frac{\sum_{k \in s_i} d_{ik}}{\sum_{i=1}^3 \sum_{k \in s_i} d_{ik}}; \quad i = 1, 2, 3$$

Lectura de base de la encuesta

```
library(haven)
library(tidyverse)
library(survey)
library(srvyr)
library(printr)
encuesta_anual <- readRDS("Data/base_anual.rds")
upm_mes <- encuesta_anual %>% distinct(trimestre, mes_trimestre, upm)

upm_mes %>% group_by(trimestre, mes_trimestre) %>%
  count(name = "n_upm") %>%
  pivot_wider(
    names_from = mes_trimestre,
    values_from = n_upm,
    values_fill = 0 # Llena con 0 si falta alguna combinación
  ) %>% mutate(Total = M1 + M2 + M3) %>% data.frame()
```

Tabla de conteo de UPM por mes y trimestre

trimestre	M1	M2	M3	Total
T1	85	84	90	259
T2	78	82	90	250
T3	69	89	92	250
T4	76	89	92	257

trimestre_1	trimestre_2	traslape
T1	T2	56
T2	T3	47
T3	T4	54

Conteo de hogares por mes y trimestre

```
hogar_mes <- encuesta_anual %>%  
  distinct(trimestre, mes_trimestre, upm, id_hogar)  
  
hogar_mes %>% group_by(trimestre, mes_trimestre) %>%  
  count(name = "n_hogar") %>%  
  pivot_wider(  
    names_from = mes_trimestre,  
    values_from = n_hogar,  
    values_fill = 0  
  ) %>% mutate(Total = M1 + M2 + M3) %>% data.frame()
```

Tabla de conteo de hogares por mes y trimestre

trimestre	M1	M2	M3	Total
T1	3607	3605	3857	11069
T2	3236	3338	3683	10257
T3	2910	3684	3793	10387
T4	3274	3675	3939	10888

Conteo de personas por mes y trimestre

```
encuesta_anual %>% group_by(trimestre, mes_trimestre) %>%  
  count(name = "n_pers") %>%  
  pivot_wider(  
    names_from = mes_trimestre,  
    values_from = n_pers,  
    values_fill = 0  
  ) %>% mutate(Total = M1 + M2 + M3) %>% data.frame()
```

Tabla de conteo de personas por mes y trimestre

trimestre	M1	M2	M3	Total
T1	10964	11101	11638	33703
T2	8892	9243	10331	28466
T3	7617	9857	10416	27890
T4	9541	9866	10965	30372

Factor de expansión ajustado por mes

```
ponderador_mes <- encuesta_anual %>%  
  group_by(trimestre, mes_trimestre) %>%  
  summarise(num = sum(fep), .groups = "drop") %>%  
  group_by(trimestre) %>%  
  mutate(den = sum(num),  
         ai = num / den)  
ponderador_mes %>% data.frame()  
  
encuesta_anual_ajus <- inner_join(encuesta_anual,  
  ponderador_mes %>% select(-num, -den),  
  by = c("trimestre", "mes_trimestre")) %>%  
  mutate(dk_mes = fep*ai)
```


Factor de expansión ajustado por mes

trimestre	mes_trimestre	num	den	ai
T1	M1	1914195	5703432	0.3356216
T1	M2	1917158	5703432	0.3361412
T1	M3	1872079	5703432	0.3282372
T2	M1	1702800	5144740	0.3309788
T2	M2	1745498	5144740	0.3392782
T2	M3	1696442	5144740	0.3297430
T3	M1	1553076	4897294	0.3171293
T3	M2	1671661	4897294	0.3413439
T3	M3	1672557	4897294	0.3415268
T4	M1	1830591	5181044	0.3533248
T4	M2	1633137	5181044	0.3152138
T4	M3	1717317	5181044	0.3314615

Total de hogares estimado por trimestre

```
encuesta_anual_ajus %>%  
  filter(id_pers == "1") %>% group_by(trimestre) %>%  
  summarise( tot_hog_dk = sum(dk_mes)) %>% data.frame()
```

trimestre	tot_hog_dk
T1	611520.0
T2	556536.4
T3	537317.3
T4	557072.4

Total de hogares estimado por mes

```
encuesta_anual_ajus %>%  
  filter(id_pers == "1") %>% group_by(trimestre, mes_trimestre) %>%  
  summarise(tot_hog_fep = sum(fep)) %>% data.frame()
```

trimestre	mes_trimestre	tot_hog_fep
T1	M1	611520.0
T1	M2	611520.0
T1	M3	611520.0
T2	M1	551018.3
T2	M2	559257.4
T2	M3	559275.5
T3	M1	533665.7
T3	M2	540305.9
T3	M3	537721.2
T4	M1	558942.3
T4	M2	561541.3
T4	M3	550829.2

Total de personas estimado por trimestre

```
encuesta_anual_ajus %>%  
  group_by(trimestre) %>%  
  summarise( tot_pers_dk = sum(dk_mes)) %>% data.frame()
```

trimestre	tot_pers_dk
T1	1901367
T2	1715190
T3	1634360
T4	1730805

Total de personas estimado por mes

```
encuesta_anual_ajus %>%  
  group_by(trimestre, mes_trimestre) %>%  
  summarise(tot_pers_fep = sum(fep)) %>% data.frame()
```

trimestre	mes_trimestre	tot_pers_fep
T1	M1	1914195
T1	M2	1917158
T1	M3	1872079
T2	M1	1702800
T2	M2	1745498
T2	M3	1696442
T3	M1	1553076
T3	M2	1671661
T3	M3	1672557
T4	M1	1830591
T4	M2	1633137
T4	M3	1717317

Total de personas estado de ocupación estimado por trimestre

```
encuesta_anual_ajus %>%  
  filter(!is.na(trabajo)) %>%  
  group_by(trimestre, trabajo) %>%  
  summarise(tot_empleo_dk = sum(dk_mes)) %>%  
  pivot_wider(  
    names_from = trabajo,  
    values_from = tot_empleo_dk,  
    values_fill = 0  
  ) %>% data.frame()
```

trimestre	Desocupado	Inactivo	Ocupado
T1	274560.2	134823.0	822042.1
T2	262841.8	134608.9	714220.4
T3	260946.4	131781.1	679790.5
T4	259658.8	121096.9	747262.1

Total de personas estado de ocupación estimado por trimestre

```
encuesta_anual_ajus %>%  
  filter(!is.na(trabajo)) %>%  
  mutate(  
    yk = ifelse(trabajo == "Desocupado", 1, 0),  
    zk = ifelse(trabajo != "Inactivo", 1, 0)  
  ) %>% group_by(trimestre) %>%  
  summarise(tot_yk = sum(dk_mes*yk),  
            tot_zk = sum(dk_mes*zk)) %>%  
  mutate(theta = tot_yk/tot_zk)%>% data.frame()
```

trimestre	tot_yk	tot_zk	theta
T1	274560.2	1096602.3	0.2503735
T2	262841.8	977062.1	0.2690123
T3	260946.4	940736.8	0.2773851
T4	259658.8	1006920.9	0.2578741

Agregación anual

- Para 12 meses, la estimación se expresa como:

$$\hat{t}_y = \sum_{s_1 \cup \dots \cup s_{12}} d_k^+ y_k \propto \sum_{i=1}^{12} \sum_{s_i} d_{ik} y_k$$

- Los pesos anuales ajustados se definen como:

$$d_{ik}^+ = b_i \cdot d_{ik}; \quad k \in s_i$$

donde:

$$b_i = \frac{\sum_{k \in s_i} d_{ik}}{\sum_{i=1}^{12} \sum_{k \in s_i} d_{ik}}; \quad i = 1, \dots, 12$$

Factor de expansión ajustado por año

```
ponderador_anos <- encuesta_anual %>%  
  group_by(trimestre, mes_trimestre) %>%  
  summarise(num = sum(fep), .groups = "drop") %>%  
  mutate(den = sum(num),  
         bi = num / den)  
ponderador_anos %>% data.frame()  
  
encuesta_anual_ajus2 <- inner_join(encuesta_anual,  
                                   ponderador_anos %>% select(-num, -den),  
                                   by = c("trimestre", "mes_trimestre")) %>%  
  mutate(dk_anual = fep*bi)
```

Factor de expansión ajustado de forma anual

trimestre	mes_trimestre	num	den	bi
T1	M1	1914195	20926510	0.0914722
T1	M2	1917158	20926510	0.0916139
T1	M3	1872079	20926510	0.0894597
T2	M1	1702800	20926510	0.0813705
T2	M2	1745498	20926510	0.0834108
T2	M3	1696442	20926510	0.0810667
T3	M1	1553076	20926510	0.0742157
T3	M2	1671661	20926510	0.0798825
T3	M3	1672557	20926510	0.0799253
T4	M1	1830591	20926510	0.0874771
T4	M2	1633137	20926510	0.0780415
T4	M3	1717317	20926510	0.0820642

Total de hogares estimado anual

```
encuesta_anual_ajus2 %>% ungroup() %>%  
  filter(id_pers == "1") %>%  
  summarise(tot_hog_dk = sum(dk_anual)) %>% data.frame()
```

tot_hog_dk

567156.9

```
encuesta_anual_ajus2 %>% ungroup() %>%  
  summarise( tot_pers_dk = sum(dk_anual)) %>% data.frame()
```

tot_pers_dk

1750882

Total de personas estado de ocupación estimado de forma anual

```
encuesta_anual_ajus2 %>%  
  filter(!is.na(trabajo)) %>%  
  group_by(trabajo) %>%  
  summarise(tot_empleo_dk = sum(dk_anual)) %>%  
  pivot_wider(  
    names_from = trabajo,  
    values_from = tot_empleo_dk,  
    values_fill = 0  
  ) %>% data.frame()
```

Desocupado	Inactivo	Ocupado
264804	130660.2	743729.9

Total de personas estado de ocupación estimado de forma anual

```
encuesta_anual_ajus2 %>% ungroup() %>%  
  filter(!is.na(trabajo)) %>%  
  mutate(  
    yk = ifelse(trabajo == "Desocupado", 1, 0),  
    zk = ifelse(trabajo != "Inactivo", 1, 0)  
  ) %>%  
  summarise(tot_yk = sum(dk_anual*yk),  
            tot_zk = sum(dk_anual*zk)) %>%  
  mutate(theta = tot_yk/tot_zk)%>% data.frame()
```

tot_yk	tot_zk	theta
264804	1008534	0.2625633

Coherencia de los nuevos factores

- Los nuevos pesos deben sumar coherentemente a la población objetivo:

Trimestral:

$$\sum_{k \in s^3} d_{ik}^+ = \sum_{i=1}^3 \sum_{k \in s_i} a_i d_{ik} \approx N$$

Anual:

$$\sum_{k \in s^{12}} d_{ik}^+ = \sum_{i=1}^{12} \sum_{k \in s_i} b_i d_{ik} \approx N$$

Recomendaciones prácticas

- ▶ Verificar que las sumas de los pesos sean coherentes no solo a nivel nacional, sino también en:
 - ▶ Ciudades principales
 - ▶ Áreas urbano/rural
 - ▶ Provincias
 - ▶ Grupos de sexo, edad, etc.
- ▶ Tras ajustar los pesos agregados, realizar una **recalibración** sobre las mismas variables utilizadas para calibrar los pesos mensuales.

Consideraciones de calibración

- ▶ Ante la falta de proyecciones trimestrales o anuales, se puede:
 - ▶ Usar el mes intermedio como referencia, o
 - ▶ Promediar los totales poblacionales de los meses incluidos.
- ▶ Este ajuste final en los pesos suele ser **mínimo** y **no altera significativamente** la estructura de los pesos originales ya calibrados.

Inclusión de todas las viviendas

- ▶ Las bases agregadas deben incluir:
 - ▶ Todas las viviendas encuestadas durante el periodo.
 - ▶ Todas las mediciones de las viviendas que participaron en más de una ocasión, especialmente bajo esquemas rotativos.
- ▶ La agregación **no distingue** entre viviendas con una o múltiples observaciones.

Esquema rotativo 2(2)2

- ▶ Consideremos una encuesta mensual continua con esquema rotativo 2(2)2.
- ▶ Si las muestras mensuales son independientes:
 - ▶ Las agregaciones trimestrales **no incluirán viviendas repetidas**.
 - ▶ Aún así, los pesos deben ajustarse en cada mes, **sin distinguir si hay repeticiones**.

Estimación del total trimestral

El estimador del total trimestral se expresa como:

$$\hat{t}_y = \sum_{s_1} d_{1k}^+ y_k + \sum_{s_2} d_{2k}^+ y_k + \sum_{s_3} d_{3k}^+ y_k = \hat{t}_y^1 + \hat{t}_y^2 + \hat{t}_y^3$$

► Donde: $d_{ik}^+ = a_i \cdot d_{ik}$

La varianza se calcula como suma de varianzas mensuales (muestreo independiente):

$$Var(\hat{t}_y) = Var(\hat{t}_y^1) + Var(\hat{t}_y^2) + Var(\hat{t}_y^3)$$

Población para calibración trimestral

```
library(openxlsx)
poblacion <- read.xlsx("Imagenes/15_cap/muestra_anual.xlsx") %>%
  transmute(estrato = paste0("hh_", dam, "_", area),
            N_hogares)
poblacion
```

estrato	N_hogares
hh_05_Urbana	79997.6
hh_05_Rural	5229.6
hh_06_Urbana	29432.8
hh_06_Rural	9153.6
hh_07_Urbana	37088.8
hh_07_Rural	9615.2
hh_08_Urbana	66091.2
hh_08_Rural	7568.8
hh_09_Urbana	34839.2
hh_09_Rural	10612.0
hh_10_Urbana	26225.6
hh_10_Rural	8456.0
hh_13_Urbana	279815.2
hh_13_Rural	7392.8

Modelo para calibración trimestral

```
total_pob <- setNames(round(poblacion$N_hogares),  
                      paste0(poblacion$estrato))  
  
formula_calb <-  
  as.formula(paste0("~ 0 + ", paste0(names(total_pob),  
                                     collapse = " + ")))  
formula_calb
```

```
~0 + hh_05_Urbana + hh_05_Rural + hh_06_Urbana + hh_06_Rural +  
    hh_07_Urbana + hh_07_Rural + hh_08_Urbana + hh_08_Rural +  
    hh_09_Urbana + hh_09_Rural + hh_10_Urbana + hh_10_Rural +  
    hh_13_Urbana + hh_13_Rural
```

Base de hogares por trimestre

```
base_hog <- encuesta_anual_ajus %>%  
  filter(id_pers == 1) %>%  
  distinct(trimestre, mes_trimestre, estrato, dam, upm,  
           id_hogar, dk_mes) %>%  
  mutate(estrato_hh = estrato) %>%  
  fastDummies::dummy_columns(select_columns = "estrato_hh",  
                             remove_selected_columns = TRUE) %>%  
  rename_with(.fn = ~ gsub("^estrato_", "", .x),  
             .cols = starts_with("estrato_hh"))
```

Definición del diseño muestral y calibración para el trimestre 1

```
df_T1 <- base_hog %>% filter(trimestre == "T1")
design_T1 <- svydesign(
  id = ~ upm,
  strata = ~ estrato,
  weights = ~ dk_mes,
  data = df_T1,
  nest = TRUE
)

calibrado_T1 <- calibrate(design = design_T1,
                        formula = formula_calb,
                        population = total_pob) %>%
  as_survey()
```

Pesos calibrados por hogar para el trimestre 1

► Pesos antes de la calibración

```
summary(weights(design_T1))
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.127272	26.16348	42.63325	55.24618	70.64404	2719.007

► Pesos después de la calibración

```
summary(weights(calibrado_T1))
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.127272	26.16348	42.63325	55.24618	70.64404	2719.007

Definición del diseño muestral y calibración para el trimestre 2

```
df_T2 <- base_hog %>% filter(trimestre == "T2")
design_T2 <- svydesign(
  id = ~ upm,
  strata = ~ estrato,
  weights = ~ dk_mes,
  data = df_T2,
  nest = TRUE
)

calibrado_T2 <- calibrate(design = design_T2,
  formula = formula_calb,
  population = total_pob) %>%
  as_survey()
```

Pesos calibrados por hogar para el trimestre 2

► Pesos antes de la calibración

```
summary(weights(design_T2))
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.5680677	25.69884	44.83415	59.43997	77.03896	891.501

► Pesos después de la calibración

```
summary(weights(calibrado_T2))
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.6360051	28.22766	49.09814	65.3124	84.72233	979.2263

Pesos calibrados por hogar para el trimestre 3

► Pesos antes de la calibración

```
summary(weights(design_T3))
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.62233	25.52703	43.71649	58.88409	72.05434	4188.1

► Pesos después de la calibración

```
summary(weights(calibrado_T3))
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.817597	28.95199	49.83338	67.01589	82.11514	4776.308

Pesos calibrados por hogar para el trimestre 4

► Pesos antes de la calibración

```
summary(weights(design_T4))
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.717369	24.9455	41.80012	55.83566	71.58804	1276.998

► Pesos después de la calibración

```
summary(weights(calibrado_T4))
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.950156	27.34964	45.80757	61.29297	78.66142	1407.399

Base de hogares con los pesos calibrados

```
base_hog_calib <-  
bind_rows(  
  design_T1$variables %>%  
    transmute(trimestre, mes_trimestre, upm, id_hogar,  
              dk_mes_cali = weights(calibrado_T1) ),  
  design_T2$variables %>%  
    transmute(trimestre, mes_trimestre, upm, id_hogar,  
              dk_mes_cali = weights(calibrado_T2) ),  
  design_T3$variables %>%  
    transmute(trimestre, mes_trimestre, upm, id_hogar,  
              dk_mes_cali = weights(calibrado_T3) ),  
  design_T4$variables %>%  
    transmute(trimestre, mes_trimestre, upm, id_hogar,  
              dk_mes_cali = weights(calibrado_T4) )  
)
```

Base de personas con los pesos calibrados

```
encuesta_anual_ajus_calib <-  
  inner_join(encuesta_anual_ajus, base_hog_calib)  
  
design_T1_calib <- svydesign(  
  id = ~ upm,  
  strata = ~ estrato,  
  weights = ~ dk_mes_cali,  
  data = encuesta_anual_ajus_calib %>%  
    filter(trimestre == "T1"),  
  nest = TRUE  
) %>% as_survey()
```

Estimación del total de personas en el trimestre 1

```
total_mes <- design_T1_calib %>% group_by(mes_trimestre) %>%  
  summarise(total_estimado = survey_total(vartype = "var"),  
            .groups = "drop")  
total_mes
```

mes_trimestre	total_estimado	total_estimado_var
M1	642445.1	5815062719
M2	644435.9	5553334911
M3	614485.9	4557547114

```
total_mes %>% summarise(  
  total_estimado_trim = sum(total_estimado),  
  total_estimado_var_trim = sqrt(sum(total_estimado_var))  
)
```

total_estimado_trim	total_estimado_var_trim
1901367	126198

Estimación del total de personas dado el estado de ocupación por mes

```
total_mes_empleo <- design_T1_calib %>%  
  filter(!is.na(trabajo)) %>%  
  group_by(mes_trimestre, trabajo ) %>%  
  summarise(total_estimado = survey_total(vartype = "var"),  
            .groups = "drop")  
total_mes_empleo
```

mes_trimestre	trabajo	total_estimado	total_estimado_var
M1	Desocupado	90071.24	123934248
M1	Inactivo	42528.65	27069606
M1	Ocupado	288458.50	1197024126
M2	Desocupado	91612.49	125547963
M2	Inactivo	50025.95	39838397
M2	Ocupado	273856.85	1038587675
M3	Desocupado	92876.48	117510599
M3	Inactivo	42268.40	24450533
M3	Ocupado	259726.77	864438430

Estimación del total de personas dado el estado de ocupación por trimestre

```
total_mes_empleo %>% group_by(trabajo ) %>% summarise(  
  total_estimado_trim = sum(total_estimado),  
  total_estimado_var_trim = sqrt(sum(total_estimado_var))  
)
```

trabajo	total_estimado_trim	total_estimado_var_trim
Desocupado	274560.2	19157.056
Inactivo	134823.0	9558.166
Ocupado	822042.1	55678.095

Estimación del total anual

El estimador del total anual es:

$$\hat{t}_y = \sum_{i=1}^{12} \sum_{s_i} d_{ik}^+ y_k = \sum_{i=1}^{12} \hat{t}_y^i$$

► Donde: $d_{ik}^+ = b_i \cdot d_{ik}$

La varianza considera dependencia entre meses (UPM compartidas):

$$Var(\hat{t}_y) = \sum_{i=1}^{12} Var(\hat{t}_y^i) + 2 \sum_{i,j=1}^{12} \sum_{j < i} Cov(\hat{t}_y^i, \hat{t}_y^j)$$

Implicaciones

- ▶ En agregación anual, no se puede asumir independencia mensual.
- ▶ Las covarianzas entre estimadores mensuales **no pueden ignorarse** debido al diseño rotativo.

Agregación de encuestas con diferentes tamaños de muestra

Agregación de encuestas con diferentes tamaños de muestra

- ▶ Las encuestas pueden tener **variaciones mensuales** importantes en el tamaño de muestra.
- ▶ Eventos externos (como la pandemia por COVID-19) pueden reducir drásticamente la cobertura en ciertos periodos.
- ▶ Aun cuando se recolectan datos cada mes, la representatividad no siempre es equivalente mes a mes.
- ▶ El ajuste correcto de los **factores de expansión** es esencial cuando se agregan los datos trimestralmente.

Ejemplo de esquema trimestral

Panel / Mes	M1	M2	M3
Panel	P1	P2	P3
Viviendas	5000	4500	2500
Panel	P4	P5	P6
Viviendas	5500	5100	3000

- ▶ Mes 1: Se usan P1 y P4
- ▶ Mes 2: Se usan P2 y P5
- ▶ Mes 3: Se usan P3 y P6 (menor tamaño muestral)

Consideraciones clave

- ▶ La agregación no debe hacerse simplemente uniendo las bases.
- ▶ Se deben tener en cuenta **las diferencias en representatividad** y ajustar los pesos.
- ▶ El objetivo es preservar la representatividad trimestral sin **sobreponderar** observaciones con menor cobertura.

Agregación con encuestas de tamaños muestrales desiguales

- ▶ Cuando el tamaño muestral varía entre periodos, como en el tercer mes del ejemplo anterior, es necesario **ajustar los pesos de muestreo** para evitar sesgos en los estimadores agregados.
- ▶ Heeringa, West, y Berglund (2017) proponen que, frente a tamaños reducidos de muestra, se pueden **normalizar los factores de expansión** mediante un ajuste proporcional, tal como fue sugerido por Kish (1999).

Ajuste ponderado

El peso ajustado se define como:

$$d_{kth}^+ = \delta_{th} \times d_{kth}$$

- ▶ d_{kth} : factor de expansión del individuo k , en el estrato h , durante el mes t .
- ▶ δ_{th} : proporción del tamaño de muestra del mes t en el estrato h con respecto al total del trimestre:

$$\delta_{th} = \frac{n_{th}}{\sum_{t=1}^3 n_{th}}$$

Interpretación de δ_{th}

- ▶ Este ajuste pondera la contribución mensual de cada estrato al total trimestral.
- ▶ Puede entenderse como un factor de **promedio ponderado**, que define la combinación lineal convexa entre los pesos.

Propiedades de los pesos ajustados trimestrales

1. **Combinación lineal convexa**

$\delta_{th} > 0$ para todo t y h , y

$$\sum_{t=1}^3 \delta_{th} = 1$$

2. **Consistencia con los tamaños poblacionales estimados por dominio**

Asumiendo que s_h es la muestra del estrato h durante los tres meses:

$$\sum_{t=1}^3 \sum_{k \in s_h} d_{kth}^+ = \sum_{t=1}^3 \delta_{th} \hat{N}_h^t \approx \hat{N}_h$$

Propiedades de los pesos ajustados trimestrales

3. **Aporte proporcional al tamaño de muestra mensual**

La suma de los factores trimestrales dentro de un dominio h para el mes t es:

$$\sum_{k \in s_{th}} d_{kth}^+ = \delta_{th} \hat{N}_h^t$$

4. **Promediación equivalente en los estratos**

El mecanismo permite representar correctamente la contribución de cada mes incluso en contextos con muestras reducidas.

- Este procedimiento permite mantener la representatividad estadística aún bajo condiciones de desbalance temporal en los tamaños muestrales, con mínimos supuestos adicionales y una estructura bien definida de ponderación.

Comprobación de la estabilidad de los factores trimestrales

Una propiedad importante del esquema de ponderación trimestral es que el aporte de los factores de expansión ponderados (d_{kth}^+) debe ser proporcional al tamaño muestral en cada dominio y cada mes.

Además, se espera que la media de los factores trimestrales sea aproximadamente constante en el tiempo dentro de cada dominio específico. Esto puede expresarse matemáticamente como:

$$\frac{\sum_{k \in s_{th}} d_{kth}^+}{n_h} = \frac{\sum_{k \in s_{th}} \delta_{th} d_{kth}}{n_h} = \frac{\sum_{k \in s_{th}} d_{kth}}{\sum_h n_h} = \frac{\hat{N}_{th}}{\sum_h n_h} \cong \frac{\hat{N}_h}{\sum_h n_h}$$

Comprobación de la estabilidad de los factores trimestrales (Nacianal)

```
encuesta_anual_ajus_calib %>%  
  group_by(trimestre, mes_trimestre) %>%  
  summarise(n_obs = n(), N_hat = sum(dk_mes_cali )) %>%  
  transmute(trimestre, mes_trimestre, prop = N_hat /n_obs ) %>%  
  pivot_wider(  
    names_from = mes_trimestre,  
    values_from = prop,  
    values_fill = 0  
  ) %>% data.frame()
```

Tabla de comprobación de la estabilidad de los factores trimestrales (Nacianal)

trimestre	M1	M2	M3
T1	58.59586	58.05206	52.79996
T2	69.65233	70.08785	59.38894
T3	72.79508	65.78096	62.76334
T4	74.03735	57.15997	56.69683

Comprobación de la estabilidad de los factores trimestrales (Sexo)

```
encuesta_anual_ajus_calib %>%  
  group_by(trimestre, mes_trimestre, sexo) %>%  
  summarise(n_obs = n(), N_hat = sum(dk_mes_cali )) %>%  
  transmute(trimestre, mes_trimestre, sexo,  
            prop = N_hat /n_obs ) %>%  
  pivot_wider(  
    names_from = mes_trimestre,  
    values_from = prop,  
    values_fill = 0  
  ) %>% data.frame()
```


Tabla de comprobación de la estabilidad de los factores trimestrales (Sexo)

trimestre	sexo	M1	M2	M3
T1	Hombre	59.98142	59.64793	54.64360
T1	Mujer	57.35817	56.56210	51.13077
T2	Hombre	70.54833	72.13663	60.76624
T2	Mujer	68.83408	68.20375	58.12127
T3	Hombre	75.11462	67.79352	66.78517
T3	Mujer	70.66331	63.89847	59.00641
T4	Hombre	76.12937	58.70903	58.64450
T4	Mujer	72.11268	55.72720	54.88880

Comprobación de la estabilidad de los factores trimestrales (Empleo)

```
encuesta_anual_ajus_calib %>%  
  group_by(trimestre, mes_trimestre, trabajo) %>%  
  summarise(n_obs = n(), N_hat = sum(dk_mes_cali )) %>%  
  transmute(trimestre, mes_trimestre, trabajo,  
            prop = N_hat /n_obs ) %>%  
  pivot_wider(  
    names_from = trabajo,  
    values_from = prop,  
    values_fill = 0  
  ) %>% data.frame()
```

Tabla de comprobación de la estabilidad de los factores trimestrales (Empleo)

trimestre	mes_trimestre	Desocupado	Inactivo	Ocupado	NA.
T1	M1	63.07510	57.31624	61.67597	53.77378
T1	M2	59.56599	62.61070	60.45405	54.07195
T1	M3	54.89154	52.50733	55.52090	49.20776
T2	M1	72.47489	76.32604	72.40128	64.14291
T2	M2	73.69710	74.19538	74.08454	63.81842
T2	M3	60.53218	66.62812	62.56378	54.29385
T3	M1	77.36630	78.65216	79.89363	62.63263
T3	M2	66.98738	73.29007	69.60262	59.94201
T3	M3	66.36501	69.70846	66.78137	56.06229
T4	M1	78.09715	73.59295	76.72534	69.68011
T4	M2	59.74975	62.07010	61.33621	50.89144
T4	M3	61.53586	56.29414	59.70030	51.71019

Comprobación de la estabilidad de los factores trimestrales (Área - Sexo)

```
encuesta_anual_ajus_calib %>%  
  group_by(trimestre, mes_trimestre, area, sexo) %>%  
  summarise(n_obs = n(), N_hat = sum(dk_mes_cali )) %>%  
  transmute(trimestre, area, mes_trimestre, sexo,  
            prop = N_hat /n_obs ) %>%  
  pivot_wider(  
    names_from = mes_trimestre,  
    values_from = prop,  
    values_fill = 0  
  ) %>% data.frame()
```

Tabla de comprobación de la estabilidad de los factores trimestrales (Área - Sexo)

trimestre	area	sexo	M1	M2	M3
T1	Urbana	Hombre	68.81466	68.86135	61.68862
T1	Urbana	Mujer	64.96196	64.37086	56.94203
T1	Rural	Hombre	28.44863	27.18718	27.27259
T1	Rural	Mujer	26.60520	25.65746	25.13633
T2	Urbana	Hombre	82.03570	82.45052	68.38925
T2	Urbana	Mujer	78.29341	76.01196	64.41596
T2	Rural	Hombre	32.39654	34.26462	32.03581
T2	Rural	Mujer	30.75278	33.29985	29.48426
T3	Urbana	Hombre	87.75085	76.97974	75.67633
T3	Urbana	Mujer	80.67205	71.87237	65.83444
T3	Rural	Hombre	35.00344	33.10359	32.74362
T3	Rural	Mujer	32.58159	30.51397	29.96313
T4	Urbana	Hombre	87.59540	66.00973	65.99450
T4	Urbana	Mujer	82.08842	61.44609	60.42617
T4	Rural	Hombre	33.94897	30.55830	30.26206
T4	Rural	Mujer	32.39187	29.57805	28.84757

Comprobación de la estabilidad de los factores trimestrales

Esto implica que, **sin importar el mes**, la media de los factores trimestrales será similar dentro de cada dominio. Este comportamiento favorece la agregación, dado que el **peso del mes con mayor tamaño de muestra influye más en el estimador agregado**, manteniendo coherencia en la estructura de ponderación.

Por tanto, para dominios con amplia muestra, como las cinco ciudades principales, la agregación trimestral tiende a producir estimaciones cercanas al promedio de los valores mensuales.

Extensión a la agregación anual

La metodología puede extenderse a la agregación de 12 meses. Para ello, se recomienda ajustar los pesos originales d_{kth} mediante un factor de normalización definido como:

$$d_{kth}^+ = \delta_{th} \cdot d_{kth}$$

donde δ_{th} representa el **porcentaje de individuos observados en el mes t para el estrato h** , calculado como:

$$\delta_{th} = \frac{n_{th}}{\sum_{t=1}^{12} n_{th}}$$

Este esquema conserva la proporcionalidad del aporte muestral en la estimación agregada anual.

Efecto del tipo de encuesta en la eficiencia de los indicadores

Efecto del tipo de encuesta en la eficiencia de los indicadores

Lograr una estimación adecuada del error de muestreo en comparaciones de múltiples periodos, con o sin agregación, es una tarea clave del investigador. Según el tipo de parámetro, la naturaleza del error cambia, así como el tamaño de muestra requerido. A continuación se ilustra esto con el caso de los **cambios netos**.

Cambios netos

Considere el cambio neto en la media de una variable de interés y entre dos periodos, t_2 y t_1 :

$$\Delta = \bar{y}_2 - \bar{y}_1$$

Este parámetro se estima de forma aproximadamente insesgada como:

$$\hat{\Delta} = \hat{\bar{y}}_2 - \hat{\bar{y}}_1 = \frac{\sum_{k \in s_2} \frac{y_k}{\pi_k}}{\sum_{k \in s_2} \frac{1}{\pi_k}} - \frac{\sum_{k \in s_1} \frac{y_k}{\pi_k}}{\sum_{k \in s_1} \frac{1}{\pi_k}}$$

donde s_1 y s_2 son las muestras seleccionadas en los dos periodos, y π_k es la probabilidad de inclusión del individuo k .

Cambios netos

La varianza de este estimador está dada por:

$$Var(\hat{\Delta}) = Var(\hat{y}_2) + Var(\hat{y}_1) - 2Cov(\hat{y}_2, \hat{y}_1)$$

La covarianza puede expresarse como:

$$2Cov(\hat{y}_2, \hat{y}_1) = 2\sqrt{Var(\hat{y}_2)}\sqrt{Var(\hat{y}_1)}\sqrt{T_2}\sqrt{T_1}R_{12}$$

donde T_1 y T_2 son los porcentajes de muestra común entre periodos, y R_{12} es la correlación de la variable x en los dos levantamientos.

Cambios netos

Asumiendo varianzas iguales $Var(\hat{y}_1) = Var(\hat{y}_2) = Var(\hat{y})$ y traslape común $T_1 = T_2 = T$, la varianza se simplifica:

$$Var(\hat{\Delta}) = 2Var(\hat{y})(1 - TR_{12})$$

Estimación del cambio neto del ingreso.

```
design_T1_T2_calib <- svydesign(  
  id = ~ upm,  
  strata = ~ estrato, weights = ~ dk_mes_cali,  
  data = encuesta_anual_ajus_calib %>%  
    filter(trimestre %in% c("T2", "T1")),  
  nest = TRUE ) %>% as_survey()  
  
(media_ingresos <- svyby(~ingreso, ~trimestre,  
  design = design_T1_T2_calib,  
  svymean, vartype = "var",  
  covmat = TRUE))
```

		trimestre	ingreso	var
T1	T1		373837.0	416572773
T2	T2		385419.4	398553562

Estimación de la cambio $\Delta = \bar{y}_2 - \bar{y}_1$ ingreso.

```
(vcov_mat <- attr(media_ingresos, "var"))
```

416572773	93803106
93803106	398553562

```
(delta <- svycontrast(media_ingresos,  
                        quote(`T2` - `T1`)))
```

	nlcon	SE
contrast	11582	25050

Impacto del tipo de encuesta

Según Kish (2004), la varianza del estimador de cambio varía según el diseño:

► **Encuesta repetida** ($T = 0$):

$$Var(\hat{\Delta}) = 2Var(\hat{y})$$

► **Encuesta de panel** ($T = 1, R_{12} > 0$):

$$Var(\hat{\Delta}) = 2Var(\hat{y})(1 - R_{12})$$

► **Encuesta rotativa** ($T \neq 0, R_{12} > 0$):

$$Var(\hat{\Delta}) = 2Var(\hat{y})(1 - TR_{12})$$

Si $R_{12} > 0$, se cumple:

$$2Var(\hat{y})(1 - R_{12}) < 2Var(\hat{y})(1 - TR_{12}) < 2Var(\hat{y})$$

Observación

El diseño de panel requiere un tamaño de muestra **menor** para estimar cambios netos, en comparación con una encuesta repetida sin traslape. El diseño rotativo representa un punto intermedio en eficiencia.

Promedio trimestral

Suponga que se desea estimar el promedio trimestral de la variable de interés y , a partir de una encuesta continua mensual para tres periodos: t_3 , t_2 y t_1 :

$$\Theta = \frac{\bar{y}_3 + \bar{y}_2 + \bar{y}_1}{3}$$

Un estimador insesgado de este parámetro es:

$$\hat{\Theta} = \frac{1}{3} (\hat{\bar{y}}_3 + \hat{\bar{y}}_2 + \hat{\bar{y}}_1) = \frac{1}{3} \left(\frac{\sum_{k \in s_3} \frac{y_k}{\pi_k}}{\sum_{k \in s_3} \frac{1}{\pi_k}} + \frac{\sum_{k \in s_2} \frac{y_k}{\pi_k}}{\sum_{k \in s_2} \frac{1}{\pi_k}} + \frac{\sum_{k \in s_1} \frac{y_k}{\pi_k}}{\sum_{k \in s_1} \frac{1}{\pi_k}} \right)$$

donde s_1 , s_2 y s_3 son las muestras correspondientes a cada mes, y π_k la probabilidad de inclusión del elemento k .

Varianza del estimador

La varianza del estimador está dada por:

$$\begin{aligned} Var(\hat{\Theta}) = \frac{1}{9} [& Var(\hat{y}_3) + Var(\hat{y}_2) + Var(\hat{y}_1) + \\ & 2Cov(\hat{y}_3, \hat{y}_2) + 2Cov(\hat{y}_3, \hat{y}_1) + 2Cov(\hat{y}_2, \hat{y}_1)] \end{aligned}$$

Asumiendo varianzas homogéneas y traslape común por diseño, además de errores débilmente estacionarios, se simplifica a:

$$Var(\hat{\Theta}) = \frac{1}{9} Var(\hat{y}) [3 + 6TR]$$

donde R es la correlación constante entre las estimaciones en diferentes meses y T es el porcentaje de traslape de muestra entre meses.

Estimación del promedio para 3 trimestre

```
design_T1_T2_T3_calib <- svydesign(  
  id = ~ upm,  
  strata = ~ estrato,  
  weights = ~ dk_mes_cali,  
  data = encuesta_anual_ajus_calib %>%  
    filter(trimestre %in% c("T2", "T1", "T3")),  
  nest = TRUE  
) %>% as_survey()  
  
(  
  media_ingresos <- svyby(  
    ~ ingreso,  
    ~ trimestre,  
    design = design_T1_T2_T3_calib,  
    svymean,  
    vartype = "var",  
    covmat = TRUE  
  )  
)
```

Tabla de estimación del promedio para 3 trimestre

		trimestre	ingreso	var
T1	T1		373837.0	420957961
T2	T2		385419.4	401945544
T3	T3		395961.3	422380550

```
(media_trimestral <- mean(media_ingresos$ingreso))
```

```
[1] 385072.6
```

```
vcov(media_ingresos)
```

	T1	T2	T3
T1	420957961	98018563	-15544988
T2	98018563	401945544	98823432
T3	-15544988	98823432	422380550

Tabla de estimación del promedio para 3 trimestre

```
(delta <- svycontrast(media_ingresos,  
                        quote((`T1` + `T2` + `T3`)/3)))
```

	nlcon	SE
contrast	385073	13366

Impacto del tipo de encuesta

Según el diseño muestral, la varianza cambia:

► **Encuesta repetida** ($T = 0$):

$$Var(\hat{\Theta}) = \frac{1}{3}Var(\hat{y})$$

► **Encuesta de panel** ($T = 1, R > 0$):

$$Var(\hat{\Theta}) = \frac{1}{9}Var(\hat{y})[3 + 6R]$$

► **Encuesta rotativa** ($T \neq 0, R > 0$):

$$Var(\hat{\Theta}) = \frac{1}{9}Var(\hat{y})[3 + 6TR]$$

Observación

Si $R > 0$, se cumple:

$$\frac{1}{9}Var(\hat{y})[3 + 6R] > \frac{1}{9}Var(\hat{y})[3 + 6TR] > \frac{1}{3}Var(\hat{y})$$

Lo anterior indica que un diseño de **panel requiere mayor tamaño de muestra** para promedios trimestrales que una encuesta repetida. El diseño rotativo es una alternativa intermedia.

Pruebas de hipótesis sobre indicadores agregados

Pruebas de hipótesis sobre indicadores agregados

Para evaluar si un cambio en el parámetro de interés entre dos periodos es significativo, se plantea una prueba de hipótesis. Por ejemplo, se puede evaluar si la tasa de desocupación ha disminuido entre dos trimestres o años consecutivos.

En comparaciones entre grupos en un mismo corte transversal (e.g., hombres vs. mujeres), debe tenerse en cuenta que el muestreo se realiza a nivel de UPM, y que los tamaños muestrales por grupo pueden variar aleatoriamente. Para comparaciones entre periodos (e.g., dos trimestres), es clave considerar la dependencia inducida por diseños de panel rotativo.

sistema de hipótesis

El sistema de hipótesis es:

$$H_0 : \theta_2 - \theta_1 = 0 \quad \text{vs.} \quad H_1 : \theta_2 - \theta_1 \neq 0$$

El estimador de diferencia es:

$$\hat{\Delta} = \hat{\theta}_2 - \hat{\theta}_1$$

Varianza del estimador

La varianza de este estimador se expresa como:

$$Var(\hat{\Delta}) = Var(\hat{\theta}_2) + Var(\hat{\theta}_1) - 2Cov(\hat{\theta}_1, \hat{\theta}_2)$$

Donde el término de covarianza está dado por:

$$Cov(\hat{\theta}_1, \hat{\theta}_2) = \sqrt{Var(\hat{\theta}_1)}\sqrt{Var(\hat{\theta}_2)}\sqrt{T_1}\sqrt{T_2}R_{12}$$

- *Salvo que los dos estimadores se construyan sobre subconjuntos disjuntos de UPM, el término de covarianza no será nulo.*

Escenarios comunes

1. **Muestreo independiente entre periodos:**

$$T_1 = T_2 = 0 \Rightarrow Cov(\hat{\theta}_1, \hat{\theta}_2) = 0$$

(e.g., meses sin traslape).

2. **Panel rotativo 2(2)2:**

$$T_1 = T_2 \approx 0.5 \text{ y } R_{12} \neq 0$$

$$Cov(\hat{\theta}_1, \hat{\theta}_2) = \frac{1}{2} \sqrt{Var(\hat{\theta}_1)} \sqrt{Var(\hat{\theta}_2)} R_{12}$$

Escenarios comunes

3. Comparaciones entre subgrupos en un mismo mes:

- Sin independencia (e.g., hombres vs. mujeres):

$$T_1 \neq T_2, R_{12} \neq 0$$

$$Cov(\hat{\theta}_1, \hat{\theta}_2) = \sqrt{Var(\hat{\theta}_1)}\sqrt{Var(\hat{\theta}_2)}\sqrt{T_1}\sqrt{T_2}R_{12}$$

- Con independencia (e.g., ciudades diferentes):

$$R_{12} = 0 \Rightarrow Cov(\hat{\theta}_1, \hat{\theta}_2) = 0$$

Estadístico de prueba

Una vez conocida la varianza del estimador, se construye el estadístico:

$$t = \frac{\hat{\Delta}}{\sqrt{Var(\hat{\Delta})}}$$

Este estadístico se distribuye como *t-Student* con gl grados de libertad, dados por:

$$gl = \sum_{h=1}^H (n_{Ih} - 1) = \#UPM - \#Estratos$$

Estadístico de prueba

El número de grados de libertad es clave para la inferencia. Por ejemplo:

► $t_{0.975,1} = 12.7$

► $t_{0.975,20} = 2.08$

► $t_{0.975,40} = 2.02$

► $t_{0.975,\infty} = 1.96$

En el caso de subpoblaciones, los grados de libertad se ajustan como:

$$gl_{subpoblacion} = \sum_{h=1}^H v_h (n_{Ih} - 1)$$

donde $v_h = 1$ si el estrato h contiene observaciones de la subpoblación de interés, y 0 en otro caso.

Prueba de hipótesis $\hat{\Delta} = \bar{y}_{T=2} - \bar{y}_{T=1}$

```
resultado_ttest <- svyttest(ingreso ~ trimestre,  
                             design = design_T1_T2_calib)  
resultado_ttest
```

Design-based t-test

```
data: ingreso ~ trimestre  
t = 0.46237, df = 438, p-value = 0.644  
alternative hypothesis: true difference in mean is not equal to 0  
95 percent confidence interval:  
-37651.39 60816.27  
sample estimates:  
difference in mean  
11582.44
```


¡Gracias!

Email: andres.gutierrez@cepal.org

Referencias

- Bell, Phillip. 2001. «Comparison of Alternative Labour Force Survey Estimators». *Survey Methodology* 27 (1): 53-63.
- Fuller, W. 1990. «Analysis of Repeated Surveys». *Survey Methodology* 16 (2): 167-80.
- Gurney, M., y J. Daly. 1965. «A Multivariate Approach to Estimation in Periodic Sample Surveys». En *Proceedings of the Social Statistics Section, American Statistical Association*, 242-57.
- Heeringa, Steven G., Brady T. West, y Patricia A. Berglund. 2017. *Applied survey data analysis*. Chapman y Hall CRC statistics en the social y behavioral sciences series. CRC Press.
- Kish, Leslie. 1999. «Cumulating/combining population surveys». *Survey Methodology* 25 (2): 129-38.
- . 2004. *Statistical Design for Research*. Wiley classic biblioteca edición. Wiley.
<https://www.wiley.com/en-us/Statistical+Design+for+Research-p-9780471691204>.
- Korn, Edward Lee, y Barry I. Graubard. 1999. *Analysis of health surveys*. Wiley.
- Lent, Janice, Stephen M. Miller, y Martha Duff. 1999. «Effects of Composite Weights