

# Ánalisis de las encuestas de hogares con R

## Procesamiento longitudinal de las encuestas rotativas

CEPAL - Unidad de Estadísticas Sociales

# Tabla de contenidos I

Procesamiento longitudinal de las encuestas rotativas

Análisis posibles con datos longitudinales

Análisis de flujos brutos y matrices de transición

## Procesamiento longitudinal de las encuestas rotativas

# Introducción

## **¿Por qué pensar en procesamiento longitudinal?**

- ▶ Algunas oficinas requieren estadísticas de seguimiento continuo.
- ▶ Se aprovecha el esquema rotativo para generar información longitudinal.
- ▶ Requiere estructura de ponderación específica.

## **¿Qué es una encuesta longitudinal?**

- ▶ Recolecta información sobre los mismos elementos en múltiples momentos.
- ▶ Contrastá con levantamientos transversales.
- ▶ Un ejemplo: esquema rotativo 4(1)0 permite seguimiento anual de 25 %.

## Introducción

Según Lynn (2009), una encuesta longitudinal observa los mismos elementos a lo largo del tiempo. Muchas encuestas rotativas, como las de empleo, se pueden convertir en longitudinales si se estructura adecuadamente el seguimiento.

## Estimación del cambio y la varianza

- ▶ El foco está en estimar cambios entre períodos consecutivos.
- ▶ Es necesario calcular:
  - a. Varianza del periodo 1
  - b. Varianza del periodo 2
  - c. Correlación entre ambos
- ▶ Esos elementos afectan CV y tamaño de muestra.

*Uno de los retos metodológicos clave es que las muestras no son independientes. Para estimar cambios correctamente, debemos considerar la varianza en cada ronda y la correlación entre observaciones repetidas.*

Análisis posibles con datos longitudinales

## Caracterización de transiciones individuales

- ▶ Identificación de hogares/personas que cambian de estatus.
- ▶ Análisis de las características de quienes *entrar* o *salen* de situaciones como la pobreza extrema, incluso sin cambios netos agregados.

### *Estabilidad e inestabilidad en el tiempo*

- ▶ Seguimiento prolongado permite detectar *trayectorias persistentes* o *fluctuaciones*.
- ▶ Comprensión más profunda de los factores que **explican la permanencia en condiciones como la pobreza extrema.**

# Duración, eventos e impactos

## *Caracterización de eventos y duración*

- ▶ Estudio de la *duración de estados*: cuánto tiempo se permanece desempleado, inactivo, fuera del sistema educativo, etc.
- ▶ Posibilidad de construir *indicadores de duración y persistencia*.

## *Evaluación de impactos y relaciones causales*

- ▶ Estimación del **efecto de intervenciones** o choques externos (ej. COVID-19) sobre fenómenos como la **desocupación**.

## Diseño de paneles rotativos en encuestas de hogares

- ▶ En América Latina, varias encuestas de hogares incorporan esquemas de panel rotativo que permiten la observación repetida de una misma unidad de análisis.
- ▶ Este diseño busca:
  - ▶ Capturar dinámicas intra-hogar e interpersonales a lo largo del tiempo.
  - ▶ Generar estimaciones robustas sobre cambios de estado (e.g., ocupación inactividad).
- ▶ La dimensión longitudinal se configura sobre los hogares que *respondieron efectivamente en más de un periodo*, permitiendo análisis más allá de los cortes transversales.

## Traslapes muestrales en un esquema 4(0)1

- ▶ Una encuesta con diseño **4(0)1** realiza cuatro observaciones trimestrales consecutivas por vivienda antes de su rotación definitiva.
- ▶ La rotación de paneles genera **traslapes sistemáticos** entre periodos consecutivos:
  - ▶ **T1 vs T2 → 75 %** de hogares compartidos.
  - ▶ **T1 vs T3 → 50 %**
  - ▶ **T1 vs T4 → 25 %**
  - ▶ **T1 vs T5 → 0 %** (panel completamente renovado).
- ▶ Esta propiedad escalonada permite construir **bases longitudinales de corto, mediano y largo plazo**, dependiendo del objetivo analítico.

# Traslapes muestrales en un esquema 4(0)1

Tabla 1: *Rotación de páneles para un diseño 4(0)1.*

Trimestre	Panel 1	Panel 2	Panel 3	Panel 4
T1	$a_1$	$b_1$	$c_1$	$d_1$
T2	$b_1$	$c_1$	$d_1$	$a_2$
T3	$c_1$	$d_1$	$a_2$	$b_2$
T4	$d_1$	$a_2$	$b_2$	$c_2$
T5	$a_2$	$b_2$	$c_2$	$d_2$
T6	$b_2$	$c_2$	$d_2$	$a_3$

# Construcción de esquemas longitudinales en paneles rotativos

La figura ilustra tres esquemas longitudinales que pueden derivarse del diseño rotativo:

1. **Bimestral (T1–T2):**
2. **Trimestral extendido (T1–T3):**
3. **Anual (T1–T4):**

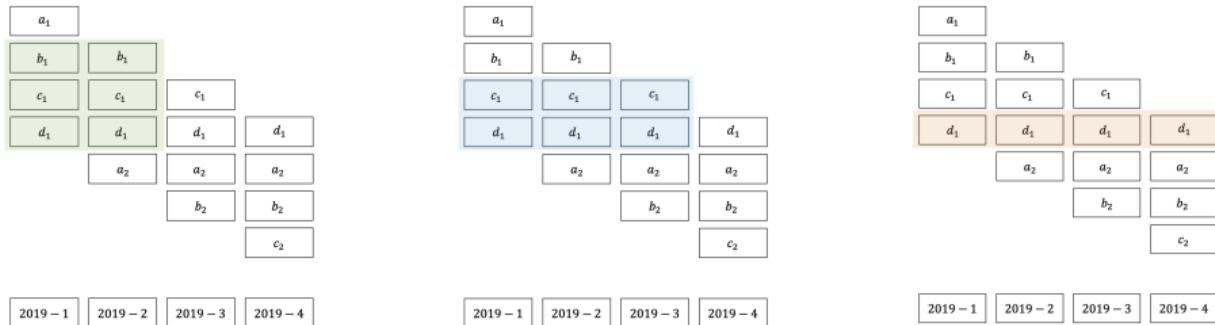


Figura 1: Tres escenarios longitudinales en un esquema rotativo 4(0)1.

## Efectos del COVID-19 en el esquema rotativo – Año 2020

- ▶ El año 2020 representó un quiebre en la operatividad normal de los levantamientos debido a la emergencia sanitaria provocada por el COVID-19.
- ▶ Las *restricciones de movilidad* obligaron a:
  - ▶ Cambiar el modo de recolección a entrevistas **telefónicas**, reduciendo la cobertura y la tasa de respuesta efectiva.
  - ▶ *Repetir el diseño muestral* del primer trimestre en el tercer trimestre, lo que alteró la lógica original de rotación 4(0)1.

## Esquema observado:

Año	Trimestre	Panel 1	Panel 2	Panel 3	Panel 4
2020	T1	$a_1$	$b_1$	$c_1$	$d_1$
	T2	$b_1$	$c_1$	$d_1$	$a_2$
	T3	$b_1$	$c_1$	$d_1$	$a_2$
	T4	$c_1$	$d_1$	$a_2$	$b_2$

### Implicaciones en el traslape muestral:

- ▶ **T2 vs T3:** 100 % de traslape.
- ▶ **T1 vs T3:** 75 % de traslape.
- ▶ **T1 vs T4:** 50 % de traslape.

## Cargue de base de datos y librerias

```
library(printr)      # Mejora la presentación de tablas en documentos RMarkdown
library(tidyverse)   # Conjunto de paquetes para manipulación y
                     # visualización de datos
library(tidyr)       # Manejo de estructuras anchas/largas
library(pROC)         # Curvas ROC
library(survey)       # Análisis de encuestas con diseño muestral complejo

# Carga de la base de datos a nivel de personas
base_personas <- readRDS(file.path(input, "base_anual.rds")) %>%
  filter(trimestre %in% c("T1", "T2")) %>%
  ungroup() # Eliminar agrupamientos previos

# Creación de la base a nivel de hogares, extrayendo variables
# clave sin duplicados

base_hogares <- base_personas %>%
  distinct(upm, trimestre, idhogar, fep)

# Visualización preliminar de la base de hogares
head(base_hogares, 10)
```

## Cargue de base de datos y librerias

upm	trimestre	id_hogar	fep
5106300296	T2	11194	151.8059
5106300296	T2	11195	261.0321
5106300296	T2	11196	190.6830
5106300296	T2	11197	216.6011
5106300296	T2	11198	120.3339
5106300296	T2	11199	309.1656
5106300296	T2	11200	207.3446
5106300296	T2	11201	151.8059
5106300296	T2	11202	290.6527
5106300296	T2	11203	290.6527

## Muestra de UPMs por trimestre

- Número de hogares por trimestre

```
base_hogares %>% group_by(trimestre) %>%
  tally(name = "hogares")
```

trimestre	hogares
T1	11069
T2	10257

- Número de UPMs únicas por trimestre

```
base_hogares %>% distinct(trimestre, upm) %>%
  group_by(trimestre) %>% tally(name = "upm")
```

trimestre	upm
T1	259
T2	250

# Número de UPMs en el traslape

Paso 1: identificar hogares que aparecen en ambos trimestres

```
# Identifica hogares que aparecen exactamente en dos trimestres
hogares_ambos <- base_hogares %>%
  group_by(id_hogar) %>%
  count() %>% filter(n == 2) %>% # Aparecen en dos trimestres
  pull(id_hogar)

# Extrae los pesos del trimestre 1 para los hogares que serán comparados
base_t1 <- base_hogares %>%
  filter(trimestre == "T1") %>% select(id_hogar, fep_t1 = fep)

# Número total de hogares con traslape en dos trimestres
length(hogares_ambos)
```

[1] 2338

## Fundamentos para la generación de bases longitudinales

- ▶ El análisis longitudinal permite observar transiciones individuales o de hogares entre estados (ocupación, pobreza, etc.) y **no es viable en todas las encuestas**, solo en aquellas con *esquemas rotativos* planificados.
- ▶ Es posible construir **tablas de transición** entre dos períodos a partir de observaciones empalmadas de la misma unidad.

## Enfoque de estimación de flujos brutos Feinberg y Stasny (1983):

Considera las diferencias entre pesos de muestreo en dos momentos como producto de la dinámica poblacional (ingresos y salidas del marco).

### Ejemplo:

Supongamos que un individuo fue clasificado como **empleado** tanto en el periodo  $t - 1$  como en el periodo  $t$ .

1. Asuma que su peso muestral en el primer periodo fue  $w_k^{t-1} = 300$  y en el segundo  $w_k^t = 305$ :

- ▶ Se asignan **300** (el mínimo entre ambos pesos) a la celda (*Empleado* → *Empleado*).
- ▶ La diferencia **5** se atribuye a la celda (*Fuera* → *Empleado*), asumiendo que provienen de entradas netas a la población de interés.

## Ejemplo:

2. Inversamente, si  $w_k^{t-1} = 305$  y  $w_k^t = 300$ :
  - ▶ Se asignan **300** a la celda (*Empleado* → *Empleado*).
  - ▶ La diferencia **5** se asigna a la celda (*Empleado* → *Fuera*), representando salidas netas del marco poblacional.

# Metodología Verma, Betti, y Ghellini (2006):

## 1. Pesos iniciales (transversales):

Basados en el diseño muestral, ajustados por:

- ▶ Probabilidad de selección por panel.
- ▶ No respuesta y cobertura.

## 2. Pesos longitudinales (dos períodos):

Ajustes aplicados:

- ▶ **Población longitudinal efectiva** (hogares presentes en ambos períodos).
- ▶ **Atrición muestral** (pérdida de casos por falta de seguimiento).
- ▶ **Calibración final** para alinear con totales poblacionales conocidos.

## Consideraciones técnicas

- ▶ El tamaño de la base longitudinal **disminuye** a medida que se incrementa el número de periodos integrados (máximo 4 en diseño 4(0)1).
- ▶ **Agrega mediciones, pero reduce unidades únicas:** más observaciones por individuo, menos individuos distintos.

## Consolidación de bases longitudinales

El primer paso en la generación de pesos longitudinales consiste en consolidar las bases de datos correspondientes a los períodos de interés. Esta integración produce bases de distintos tamaños según la cantidad de períodos combinados —por ejemplo, dos, tres o cuatro trimestres consecutivos.

*En términos generales:*

1. Cuantos menos períodos se integren, mayor será el número de unidades observacionales disponibles.
2. En el caso específico de un esquema rotativo 4(0)1, no es posible consolidar cinco períodos consecutivos, ya que la rotación garantiza traslape máximo en solo cuatro trimestres consecutivos.

# Implicaciones de la integración de paneles

La consolidación de paneles implica dos efectos clave:

- ▶ *Agregación de información:* Se repiten las observaciones de los mismos individuos en múltiples periodos, lo que enriquece el análisis longitudinal y permite estimar dinámicas de cambio.
- ▶ *Reducción de unidades observacionales:* Al requerir presencia continua en todos los periodos seleccionados, se pierde cobertura muestral frente al total de la muestra transversal en cada periodo.

Por tanto, se debe balancear el análisis de cambios individuales con la representatividad estadística que puede verse comprometida al aumentar la exigencia de continuidad en los datos.

## Creación de los pesos longitudinales iniciales

El proceso inicia con la definición de los **periodos consecutivos** que se desean combinar.

En el caso del año 2020, se deben considerar los cambios operativos causados por la pandemia:

- ▶ A partir del segundo trimestre, los levantamientos dejaron de ser presenciales.
- ▶ Esto impactó tanto la tasa de respuesta como la comparabilidad entre trimestres.

Solo se incluirán las unidades muestrales que **respondieron en todos los períodos** seleccionados.

## Determinación de pesos básicos

Los pesos longitudinales iniciales se derivan a partir de los **pesos básicos ajustados por cobertura** del primer periodo de combinación.

Ejemplos:

- ▶ Para combinar **T1 y T2 de 2020**, se parte de los pesos ajustados del **primer trimestre**.
- ▶ Para combinar **T2 y T3**, se toman los pesos del **segundo trimestre**.

Según LaRoche (2003), los pesos básicos se corrigen por la probabilidad de selección de paneles:

$$d_{1,k}^{\text{básico}} = \frac{d_{1,k}}{\Pr(\text{selección de paneles})}$$

## Consideraciones y validaciones

- ▶ En el ejemplo de **T1 y T2**, si tres paneles coinciden de cuatro posibles, entonces:  
 $\Pr(\text{selección de paneles}) = \frac{3}{4}$ .
- ▶ En **T2 y T3**, con la muestra replicada debido a la pandemia:  
 $\Pr(\text{selección de paneles}) = \frac{4}{4} = 1$ .

### Validación clave:

La suma de pesos básicos debe aproximar el tamaño poblacional:

$$\sum_{s^{(1)}} d_{1,k}^{\text{básico}} \approx N$$

Además, como propone la metodología de la *Survey of Labour and Income Dynamics* (Naud 2002; LaRoche 2003), este ajuste inicial incorpora la **probabilidad de traslape**.

## Creación de los pesos longitudinales

Paso 2: crear variable dicotómica de respuesta en ambos trimestres

```
base_t1_t2 <- base_hogares %>%
  mutate(respboth = if_else(id hogar %in% hogares_ambos, 1, 0)) %>%
  inner_join(base_t1, by = "id hogar")

head(base_t1_t2 %>% filter(respboth == 1) , 8)
```

upm	trimestre	id_hogar	fep	respboth	fep_t1
5106300297	T1	11233	131.8286	1	131.8286
5106300297	T2	11233	195.9185	1	131.8286
5106300297	T1	11234	152.4268	1	152.4268
5106300297	T2	11234	226.5307	1	152.4268
5106300297	T1	11235	148.3072	1	148.3072
5106300297	T2	11235	220.4083	1	148.3072
5106300297	T1	11236	126.3358	1	126.3358
5106300297	T2	11236	187.7552	1	126.3358

## Creación de los pesos longitudinales

Paso 3: asignar peso básico (solo a hogares que respondieron en ambos)

```
prob_panel <- 2/4
base_t1_t2 <- base_t1_t2 %>%
  mutate( fep_long = ifelse(respboth == 1, fep_t1/prob_panel,0)
  )
sum(base_t1_t2$fep_long)
```

```
[1] 1538004
```

```
base_personas %>% group_by(trimestre) %>%
  summarise(pob_estimada = sum(fep))
```

trimestre	pob_estimada
T1	5703432
T2	5144740

## Creación de los pesos longitudinales

```
head(base_t1_t2 %>%
      select(trimestre, id_hogar,
             respboth, fep_t1, fep_long), 10)
```

trimestre	id_hogar	respboth	fep_t1	fep_long
T1	11233	1	131.8286	263.6572
T2	11233	1	131.8286	263.6572
T1	11234	1	152.4268	304.8537
T2	11234	1	152.4268	304.8537
T1	11235	1	148.3072	296.6144
T2	11235	1	148.3072	296.6144
T1	11236	1	126.3358	252.6715
T2	11236	1	126.3358	252.6715
T1	11237	1	138.6947	277.3894
T2	11237	1	138.6947	277.3894

## Ajuste por ausencia de respuesta

- ▶ Sobre los **pesos básicos**, se debe realizar un ajuste por **no respuesta**.
- ▶ Este ajuste debe basarse en:
  - ▶ Covariables auxiliares disponibles en el marco de muestreo.
  - ▶ Registros administrativos o rondas previas.
- ▶ Las unidades que no respondieron deben ser excluidas:

$$d_{1,k}^{bsico} = 0, \quad \forall k \notin s_r^{(1)}$$

donde  $s_r^{(1)}$  representa el conjunto de respondientes efectivos del primer periodo.

## Modelo de propensión de respuesta

- Se modela la **probabilidad de respuesta**:

$$\phi_{1,k} = \Pr(D_{1,k} = 1 \mid I_{1,k} = 1) = f(\mathbf{x}_1, \beta)$$

donde  $D_{1,k}$  indica si la persona del hogar respondió la encuesta y  $I_{1,k}$  si la persona pertenece a la muestra del primer periodo.

- Comúnmente se usa un **modelo logístico**:

$$\hat{\phi}_{1,k} = \frac{\exp(\mathbf{x}'_1 \hat{\beta})}{1 + \exp(\mathbf{x}'_1 \hat{\beta})}$$

- Requiere que  $\mathbf{x}_1$  esté disponible para todos los seleccionados (respondan o no).

## Ajuste final del peso inicial

- ▶ Se ajustan los pesos por el inverso de la probabilidad estimada:

$$d_{1,k}^{inicial} = \frac{d_{1,k}^{bsico}}{\hat{\phi}_{1,k}}$$

- ▶ En ausencia de información auxiliar: Usar la **tasa media de respuesta** como imputación para  $\hat{\phi}_{1,k}$ .
- ▶ Si la unidad es nueva en el panel: Imputar el **peso del hogar** al que pertenece.
- ▶ Verificar:
  - ▶ **Balanceo** entre respondientes y no respondientes.
  - ▶ **Soporte común** entre distribuciones de propensión (evitar extremos 0 y 1).

# Creación de los pesos longitudinales

*Paso 4: Identificar las personas en los hogares respondieron en ambos trimestres*

Para construir una base de datos longitudinal consistente, es necesario identificar a las personas que respondieron en ambos trimestres (T1 y T2), dentro de los hogares previamente identificados como comunes.

```
base_personas_t1_t2 <- base_personas %>%
  filter(id_hogar %in% hogares_ambos) %>%
  mutate(id_llave = paste0(id_hogar, id_pers))
```

```
hogares_personas <- base_personas_t1_t2 %>%
  group_by(id_llave) %>% count() %>%
  filter(n == 2) %>% pull(id_llave)
```

## Creación de los pesos longitudinales

Identificando las personas que respondieron en los dos trimestres.

```
base_personas_t1_t2 <- base_personas_t1_t2 %>%
  mutate(respboth_per = if_else(id_llave %in% hogares_personas, 1, 0))

base_personas_t1_t2 <- base_personas_t1_t2 %>%
  inner_join(base_t1_t2 %>%
              select(id Hogar, fep_long, trimestre),
            by = c("id Hogar", "trimestre"))
```

## Ajuste por falta de respuesta de personas

### *Paso 5: estimar modelo logístico de probabilidad de respuesta*

Se utiliza un modelo de regresión logística para estimar la probabilidad de que una persona haya respondido en ambos trimestres, en función de características observables.

```
modelo_logit <- glm(respboth_per ~ sexo+ ingreso,  
                      data = base_personas_t1_t2,  
                      family = binomial(link = "logit"))  
# summary(modelo_logit)
```

## Estimación de la probabilidad de respuesta

Paso 6: predecir la probabilidad de respuesta.

Se predicen las probabilidades individuales de respuesta, las cuales luego serán utilizadas para ajustar los pesos longitudinales.

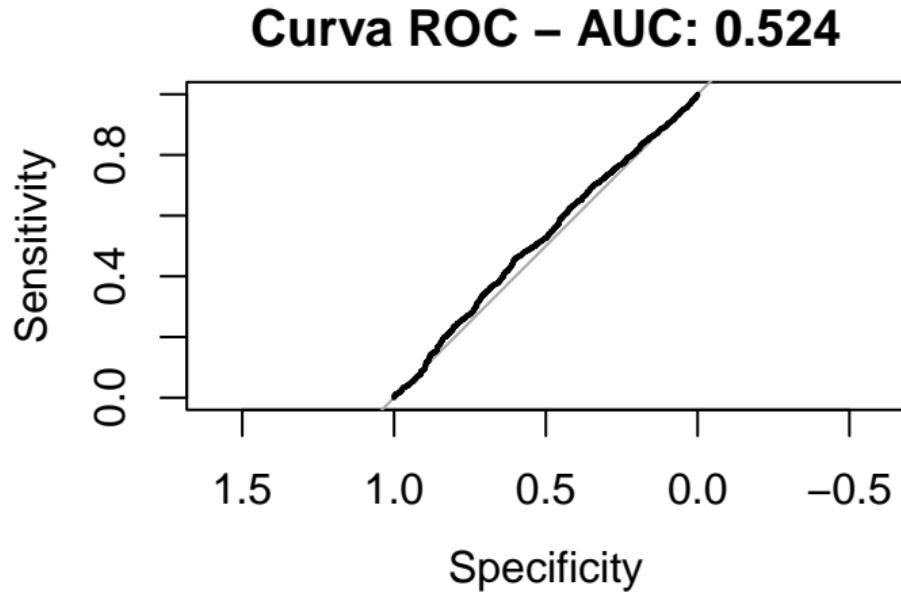
```
prob_resp = predict(modelo_logit, type = "response")
base_personas_t1_t2$prob_resp = prob_resp

roc_obj <- roc(base_personas_t1_t2$respboth_per,
                 base_personas_t1_t2$prob_resp)
```

## Evaluación del modelo: Curva ROC

La curva ROC permite verificar la capacidad predictiva del modelo de propensión. El área bajo la curva (AUC) debe acercarse a 1 para un buen ajuste.

```
plot(roc_obj, main = paste("Curva ROC - AUC:", round(auc(roc_obj), 3)))
```



## Ajuste del peso longitudinal inicial

*Paso 7: ajustar el peso longitudinal inicial usando el inverso de la probabilidad*

El factor de expansión longitudinal se ajusta dividiendo el peso original entre la probabilidad de respuesta estimada.

```
base_personas_t1_t2 <- base_personas_t1_t2 %>%
  mutate(fep_aj = fep_long / prob_resp)
```

## Visualización del ajuste de pesos

Se revisan los valores originales y ajustados para verificar consistencia y posibles extremos.

```
head(base_personas_t1_t2 %>% select(id_hogar, id_pers ,trimestre,  
fep_long, fep_aj))
```

id_hogar	id_pers	trimestre	fep_long	fep_aj
11233	1	T1	263.6572	278.1455
11233	2	T1	263.6572	279.4626
11233	3	T1	263.6572	278.1455
11233	4	T1	263.6572	278.1455
11234	1	T1	304.8537	323.0531
11234	2	T1	304.8537	321.5365

## Definición de la población longitudinal

- ▶ La población longitudinal está compuesta por las personas que permanecen en la población objetivo entre dos periodos consecutivos.
- ▶ Ejemplo: en una encuesta del 2020, la población longitudinal del primer semestre incluye personas presentes tanto en el primer como en el segundo trimestre.

## Exclusión por salida de la población

- ▶ Entre los dos periodos, pueden haber personas que salen de la población objetivo (muerte, migración, institucionalización, etc.).
- ▶ Estas personas no deben formar parte del análisis longitudinal, aunque sí están en la población objetivo del segundo periodo.

## Muestra longitudinal y asignación de pesos

- La muestra longitudinal se define como:

$$s^{(2)} = s^1 \cap s^2$$

$$\begin{array}{r} \overline{\phantom{0}} \\ 0 \qquad 1 \\ \hline 719 \quad 12814 \\ \hline \end{array}$$

- El peso longitudinal inicial se transfiere directamente:

$$d_k^{\text{final}} = d_{1,k}^{\text{inicial}}$$

**Se recomienda calibrar con totales auxiliares si están disponibles.**

## Ausencia de respuesta y atrición

- ▶ No todas las unidades responden en ambos periodos.
- ▶ Se identifican tres subconjuntos:
  - a. Respondieron en t1 pero no en t2.
  - b. No respondieron en t1 pero sí en t2.
  - c. No respondieron en ninguno.
- ▶ Solo se mantienen las unidades que respondieron en ambos.

## Asignación de peso nulo a los no respondientes

- Para unidades que no están en ambos periodos:

$$d_{2,k}^{inicial} = \begin{cases} d_{1,k}^{inicial}, & \text{si } k \in s_r^{(2)} \\ 0, & \text{si } k \notin s_r^{(2)} \end{cases}$$

- $s_r^{(2)}$  representa a las unidades respondientes en ambos periodos.

## Ajuste por respuesta en segundo periodo

- ▶ Se estima la probabilidad de respuesta en el segundo periodo:

$$\phi_{2,k} = Pr(D_{2,k} = 1 \mid I_{2,k} = 1) = f(\mathbf{x}_2, \beta)$$

- ▶ Utilizar covariables auxiliares  $\mathbf{x}_2$  disponibles para todos los seleccionados.

## Peso longitudinal ajustado

- ▶ El ajuste final del peso longitudinal se hace mediante el inverso de la probabilidad de respuesta en el segundo periodo:

$$d_{2,k}^{longitudinal} = \frac{d_{2,k}^{inicial}}{\hat{\phi}_{2,k}}$$

- ▶ Este peso se utiliza para representar adecuadamente la población longitudinal.

## Calibración de los pesos longitudinales

- ▶ Tras el ajuste por no respuesta, se recomienda calibrar los pesos.
- ▶ La calibración busca alinear los pesos con totales poblacionales conocidos:
  - ▶ Proyecciones nacionales.
  - ▶ Distribución por sexo, edad, región, área, etc.

## Restricción de calibración

- ▶ Las variables auxiliares se denotan como  $\mathbf{z}_k$ .
- ▶ Se impone la siguiente condición:

$$\sum_{s_r^{(2)}} w_{2,k}^{\text{calibrado}} \mathbf{z}_k = \sum_U \mathbf{z}_k$$

- ▶ Donde  $s_r^{(2)}$  representa a los respondientes en ambos periodos.

## Forma funcional del peso calibrado

- ▶ Los pesos calibrados se definen como:

$$w_{2,k}^{\text{calibrado}} = g_k \cdot d_{2,k}^{\text{longitudinal}}$$

- ▶  $g_k$ : factor de calibración, idealmente cercano a 1.
- ▶  $d_{2,k}^{\text{longitudinal}}$ : peso longitudinal ajustado por no respuesta.

## Consideraciones adicionales

- ▶ Las restricciones deben referirse a la población del **primer periodo**.
- ▶ La muestra longitudinal **no representa entradas posteriores**, solo a quienes estaban presentes al inicio.

# Creación de los pesos longitudinales

## Paso 8: Definir totales poblacionales conocidos

```
library(openxlsx)
poblacion <- read.xlsx("Imagenes/15_cap/muestra_anual.xlsx") %>%
  transmute(estrato = paste0("pp_", dam, "_", area),
            N_personas)
total_pob <- setNames(round(poblacion$N_personas),
                      paste0(poblacion$estrato))

formula_calb <-
  as.formula(paste0("~ 0 + ", paste0(names(total_pob), collapse = " + ")))
formula_calb

~0 + pp_05_Urbana + pp_05_Rural + pp_06_Urbana + pp_06_Rural +
  pp_07_Urbana + pp_07_Rural + pp_08_Urbana + pp_08_Rural +
  pp_09_Urbana + pp_09_Rural + pp_10_Urbana + pp_10_Rural +
  pp_13_Urbana + pp_13_Rural
```

# Creación de los pesos longitudinales

## Organizando la base de datos

```
temp <- base_personas_ambos %>%
  distinct(upm, estrato, id_hogar, id_pers, id_llave,
          edad, sexo, area, fep_aj)

temp_pobreza <- base_personas_ambos %>%
  select(id_llave, trimestre, pobreza) %>%
  pivot_wider(names_from = trimestre, values_from = pobreza) %>%
  rename(pobreza_t1 = 'T1', pobreza_t2 = 'T2')

temp <- temp %>% mutate(estrato_pp = estrato) %>%
  fastDummies::dummy_columns(select_columns = "estrato_pp",
                               remove_selected_columns = TRUE) %>%
  rename_with(.fn = ~ gsub("^estrato_", "", .x),
             .cols = starts_with("estrato_pp"))

base_personas_trim <- inner_join(temp, temp_pobreza, by = "id_llave")
```

# Creación de los pesos longitudinales

## Paso 9: Crear diseño muestral con pesos ajustados

```
library(survey)
library(srvyr)
design_long <- svydesign(
  ids = ~upm,           # Asumiendo muestreo sin conglomerados
  strata = ~estrato,    # Variable de estrato
  data = base_personas_trim,
  weights = ~fep_aj     # Peso ajustado por no respuesta
)
```

# Calibración de los pesos longitudinales

Paso 10: Ajuste por calibración por área y sexo

```
design_cali <- calibrate(  
  design = design_long,  
  formula = formula_calb, # Efectos fijos (sin intercepto)  
  population = total_pob  
) %>% as_survey()  
summary(weights(design_long)); summary(weights(design_cali))
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
7.070486	160.8165	280.7615	349.9317	463.5673	3643.129

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
32.60571	886.2015	1448.021	1879.305	2438.336	20133.45

## Calibración de los pesos longitudinales

```
options(survey.lonely.psu="certainty")
design_cali %>% group_by(estrato) %>%
  summarise(n_upm = n_distinct(upm),
            total = survey_total(vartype = "se"))
```

estrato	n_upm	total	total_se
05_Rural	2	129331	0
05_Urbana	6	1387018	0
06_Rural	2	193438	0
06_Urbana	4	574031	0
07_Rural	2	232151	0
07_Urbana	4	646128	0
08_Rural	2	147856	0
08_Urbana	7	1163871	0
09_Rural	2	229509	0
09_Urbana	4	568429	0
10_Rural	2	182406	0
10_Urbana	4	516102	0
13_Rural	1	223695	0

## Creación de los pesos longitudinales

Paso final: Asignar pesos calibrados a la base

```
base_personas_trim$fex_cali <- weights(design_cali)
```

## Análisis de flujos brutos y matrices de transición

# Análisis de flujos brutos

¿Por qué usar datos longitudinales?

- ▶ Permiten conocer el **estado de una unidad** en distintos períodos.
- ▶ Posibilitan el análisis de **flujos brutos** entre categorías (por ejemplo, ocupación).
- ▶ Se pueden descomponer los **cambios netos** estimados en cortes transversales.

*“La principal ventaja de los datos longitudinales es la posibilidad de estimar los flujos brutos.” - Lynn (2009)*

## Análisis de flujos brutos

- ▶ Una encuesta rotativa permite estimar cuántas personas **cambiaron** de estado ocupacional.
- ▶ El análisis muestra si los “nuevos ocupados” en realidad ya lo eran en el trimestre anterior.
- ▶ A partir del seguimiento en dos trimestres, se puede saber si quienes estaban empleados, desempleados o inactivos, **mantuvieron o cambiaron** su estado.

# Importancia del diseño y la no respuesta

¿Qué puede sesgar los flujos brutos?

- ▶ La **ausencia de respuesta** raramente es aleatoria.
- ▶ Por ejemplo, los desempleados tienden a no responder más frecuentemente.

Riesgos:

- ▶ Si se **ignora la ausencia de respuesta**, los flujos brutos estarán sesgados.
- ▶ Si se corrige por no respuesta, pero se ignora el **diseño muestral**, también se introduce sesgo.

# Matrices de transición y diseños complejos

## Matrices de transición

- ▶ Permiten estudiar **cambios de estado** entre dos periodos (ej. de ocupado a desempleado).
- ▶ Se construyen con unidades **respondientes en ambos períodos**.
- ▶ Fundamentales para entender dinámicas del mercado laboral.

## Diseño complejo y flujos

- ▶ Las encuestas de hogares usan **muestreo complejo**: estratificado, multietápico, con probabilidades desiguales.
- ▶ Los **factores de expansión deben reflejar** este diseño para obtener estimaciones válidas.

## Ausencia de respuesta y sesgos

- ▶ La no respuesta puede depender del **estado ocupacional** (ej. más no respuesta entre desempleados).
- ▶ Por eso, al estimar matrices de transición, se deben usar **pesos longitudinales ajustados y calibrados**.

## Modelos de Markov y transiciones en el tiempo

- ▶ Se utilizan para modelar los **cambios entre categorías** en dos períodos consecutivos.
- ▶ Asumen que los individuos pueden transitar entre **G estados mutuamente excluyentes**.
- ▶ El interés se centra en la **estimación de flujos brutos**: cambios reales entre estados.
- ▶ Solo se observa una muestra, pero el objetivo es **inferir la matriz poblacional de transición**.

# Estructura de la matriz de transición

Sea  $X_{ij}$  la **cantidad de personas** que pasaron del estado  $i$  al  $j$ .

Tabla 14: Distribución no observable de los flujos brutos en una población.

Estado (T1/T2)	1	2	...	j	...	G
1	$X_{11}$	$X_{12}$	...	$X_{1j}$	...	$X_{1G}$
2	$X_{21}$	$X_{22}$	...	$X_{2j}$	...	$X_{2G}$
:	:	:	..	:	..	:
i	$X_{i1}$	$X_{i2}$	...	$X_{ij}$	...	$X_{iG}$
:	:	:	..	:	..	:
G	$X_{G1}$	$X_{G2}$	...	$X_{Gj}$	...	$X_{GG}$

## Supuestos y limitaciones

- ▶ Se asume que la matriz de transición es **homogénea** para la población.
- ▶ La **ausencia de respuesta no ignorable** puede sesgar la estimación si no se corrige adecuadamente.
- ▶ Es fundamental **ajustar pesos y considerar el diseño muestral**.

## Modelos de Markov con no respuesta

- ▶ Suponen que cada individuo pertenece a uno de  $G$  estados.
- ▶ Se modela el cambio de estado entre  $t - 1$  y  $t$  mediante:
  - ▶  $\eta_i$ : probabilidad de estar en el estado  $i$  en  $t - 1$ .
  - ▶  $p_{ij}$ : probabilidad de transición de  $i$  a  $j$ .

### Restricciones de los parámetros

- ▶ Suma total de probabilidades iniciales:

$$\sum_i \eta_i = 1$$

- ▶ Para cada fila  $i$  en la matriz de transición:

$$\sum_j p_{ij} = 1$$

## Clasificación según respuesta

Tres grupos de individuos:

1. Respondieron en ambos periodos  $\rightarrow$  matriz  $G \times G$ .
2. Solo respondieron en uno  $\rightarrow$  complemento fila o columna.
3. No respondieron en ninguno  $\rightarrow$  celda  $M$ .

## Distribución observable con no respuesta

T1 / T2	Formal	Informal	Desocupado	Inactivo	Comp. fila
Formal	$N_{11}$	$N_{12}$	$N_{13}$	$N_{14}$	$R_1$
Informal	$N_{21}$	$N_{22}$	$N_{23}$	$N_{24}$	$R_2$
Desocupado	$N_{31}$	$N_{32}$	$N_{33}$	$N_{34}$	$R_3$
Inactivo	$N_{41}$	$N_{42}$	$N_{43}$	$N_{44}$	$R_4$
Comp. col.	$C_1$	$C_2$	$C_3$	$C_4$	$M$

## Segunda etapa: proceso de respuesta

Cada individuo puede responder o no en cada periodo. Se introducen:

1.  $\psi(i, j)$ : probabilidad de respuesta en  $t - 1$  estando en celda  $ij$ .
2.  $\rho_{RR}(i, j)$ : probabilidad de seguir respondiendo en  $t$ .
3.  $\rho_{MM}(i, j)$ : probabilidad de seguir sin responder en  $t$ .

## Modelos reducidos de no respuesta

Para estimar los parámetros de respuesta, se proponen modelos con supuestos específicos sobre la dependencia entre la respuesta y el estado de clasificación laboral.

### **Modelo A:** Probabilidades constantes

- ▶  $\psi(i, j) = \psi$ : misma probabilidad de respuesta en  $t - 1$  para todos.
- ▶  $\rho_{RR}(i, j) = \rho_{RR}$ : misma probabilidad de mantenerse respondiente.
- ▶  $\rho_{MM}(i, j) = \rho_{MM}$ : misma probabilidad de mantenerse ausente.
- ▶ **No depende del estado laboral.**

## Modelos reducidos de no respuesta

### Modelo B: Respuesta inicial según estado

- ▶  $\psi(i, j) = \psi(i)$ : depende del estado en  $t - 1$ .
- ▶  $\rho_{RR}(i, j) = \rho_{RR}$ ,  $\rho_{MM}(i, j) = \rho_{MM}$ : constantes.
- ▶ **Distingue entre formales, informales, etc. al inicio.**

### Modelo C: Transiciones según estado en $t - 1$

- ▶  $\psi(i, j) = \psi$ : constante.
- ▶  $\rho_{RR}(i, j) = \rho_{RR}(i)$ : depende del estado en  $t - 1$ .
- ▶  $\rho_{MM}(i, j) = \rho_{MM}(i)$ .
- ▶ **Transiciones diferenciadas por estado inicial.**

## Modelos reducidos de no respuesta

**Modelo D:** Transiciones según estado en  $t$

- ▶  $\psi(i, j) = \psi$ : constante.
- ▶  $\rho_{RR}(i, j) = \rho_{RR}(j)$ ,  $\rho_{MM}(i, j) = \rho_{MM}(j)$ : dependen del estado final.
- ▶ **Transiciones diferenciadas por estado destino.**

## Estimación con pseudo-verosimilitud

- ▶ Metodología de Feinberg y Stasny (1983) extendida por Gutiérrez (2014).
- ▶ Se usa **máxima pseudo-verosimilitud** bajo muestreo complejo.
- ▶ Implementado en el paquete `surf` de R Jacob (2020).

## Estimación de matrices de transición

- ▶ Se analiza la transición laboral de **41,274 personas** en los dos primeros trimestres de 2020.
- ▶ La clasificación incluye: formal, informal, desocupado e inactivo.
- ▶ Se observa **alta no respuesta en T2** (complemento fila) → efecto pandemia.

## Muestra observada: transiciones T1-T2

Tabla 16: Distribución observada de los flujos brutos en la muestra no ponderada con ausencia de respuesta en ambos periodos.

Estado (T1/T2)	Formal	Informal	Desocupado	Inactivo	Complemento fila
Formal	11,483	718	592	1,828	451
Informal	703	2,513	495	2,769	198
Desocupado	191	181	503	794	81
Inactivo	364	641	388	15,386	382
Complemento columna	160	65	48	257	83

## Estimación poblacional (ponderada)

Tabla 17: Distribución poblacional estimada de los flujos brutos ausencia de respuesta en ambos periodos.

Estado (T1/T2)	Formal	Informal	Desocupado	Inactivo	Complemento fila
Formal	3,269,673	201,639	175,719	503,740	155,902
Informal	232,095	641,565	146,416	725,006	58,649
Desocupado	55,243	50,337	157,642	233,597	26,695
Inactivo	102,490	161,363	98,898	4,299,066	118,393
Complemento columna	47,104	26,276	19,746	100,775	25,545

Total estimado: **15,597,572**

## Comparación de modelos

Criterio: ajuste de los modelos mediante estadístico chi-cuadrado.

Tabla 18: Ajuste de los cuatro modelos.

	Modelo A	Modelo B	Modelo C	Modelo D
Grados de libertad	7	4	1	1
Valor crítico	14.07	9.49	3.84	3.84
Valor $\chi^2_{RS}$	15.6706	18.3659	0.2418	3.9137

## Modelo seleccionado: Modelo C

- ▶ Mejor ajuste:  $\chi^2 = 0.2418 <$  valor crítico.
- ▶ Implica:
  - ▶ Probabilidad de respuesta constante en  $t - 1$ .
  - ▶ Transiciones de respuesta dependen del estado en  $t - 1$ .
- ▶ **Captura correctamente el patrón observado en pandemia.**

# Estimación poblacional de los flujos brutos

## **Modelo C** Proceso no observable

El modelo C asume:

- ▶ Probabilidad de respuesta igual para todos los estados en el tiempo  $t - 1$ :  
 $\psi(i, j) = \psi$ .
- ▶ Las probabilidades de transición de respuesta dependen del estado en el tiempo  $t - 1$ :  $\rho_{RR}(i, j) = \rho_{RR}(i)$ ,  $\rho_{MM}(i, j) = \rho_{MM}(i)$ .

## Tabla de distribución poblacional estimada

Tabla 19: Distribución poblacional estimada de los flujos brutos para el proceso no observable (sin ausencia de respuesta) en ambos periodos bajo el modelo C. Los errores estándar se muestran en paréntesis.

Estado (T1/T2)	Formal	Informal	Desocupado	Inactivo
Formal	4.627.632 (102.470)	287.124 (18.979)	252.084 (20.399)	713.284 (30.358)
Informal	327.066 (33.292)	911.996 (45.645)	210.372 (16.508)	1.022.351 (50.332)
Desocupado	79.346 (10.592)	72.944 (8.858)	230.949 (21.180)	335.746 (28.023)
Inactivo	143.303 (11.192)	227.545 (15.849)	140.923 (10.550)	6.014.907 (123.559)

- ▶ La diagonal principal refleja estabilidad en el estatus laboral (p. ej. 6 millones de inactivos que permanecen como tal).
- ▶ Las celdas fuera de la diagonal permiten identificar los flujos entre estados, con sus respectivos errores estándar.

## Estimación de las matrices de transición laboral

Tabla 20: Estimación de las matrices de transición laboral en ambos periodos bajo el modelo C. Los errores estándar se muestran en paréntesis.

Estado (T1/T2)	Formal	Informal	Desocupado	Inactivo
Formal	0.787 (0.010)	0.048 (0.003)	0.042 (0.003)	<b>0.121</b> (0.004)
Informal	0.132 (0.012)	0.368 (0.013)	0.085 (0.005)	<b>0.413</b> (0.013)
Desocupado	0.110 (0.012)	0.101 (0.010)	0.321 (0.021)	<b>0.466</b> (0.025)
Inactivo	0.021 (0.001)	0.034 (0.002)	0.021 (0.001)	<b>0.921</b> (0.010)

**Transiciones laborales:** Se estima que:

- ▶ El **12,1 %** de los trabajadores formales pasaron a la **inactividad**.
- ▶ **41,3 %** de los informales e **46,6 %** de los desocupados también cayeron en la **inactividad**.
- ▶ El **92,1 %** de los inactivos se mantuvo en ese estado.

## Modelo C: Resultados clave

Tabla 21: Estimación de los demás parámetros del modelo C. Los errores estándar se muestran en paréntesis.

Trimestre I-2020	$\hat{\eta}$	$\hat{\rho}_{RR}$	$\hat{\rho}_{MM}$
Formal	0.376 (0.004)	0.963 (0.004)	0.304 (0.104)
Informal	0.158 (0.003)	0.967 (0.004)	0.000 (0.261)
Desocupado	0.046 (0.002)	0.949 (0.008)	0.000 (0.188)
Inactivo	0.418 (0.004)	0.975 (0.002)	0.017 (0.086)

## Comparación por sexo: Impacto diferencial de la pandemia

El modelo C fue ajustado por sexo, mostrando buen ajuste en ambos grupos:  
 $\chi^2_{\text{hombres}} = 0.350$  y  $\chi^2_{\text{mujeres}} = 0.470$ .

La probabilidad inicial de respuesta fue igual para ambos grupos:  $\hat{\psi} = 0,981$  (EE = 0,002).

La estimación de las probabilidades  $\hat{\rho}_{MM}$  reveló diferencias por sexo:

- ▶ **Formales:** mujeres más propensas a no responder en el segundo trimestre ( $\hat{\rho}_{MM} = 0,331$  vs. 0,253).
- ▶ **Inactivos:** hombres con mayor propensión a no responder ( $\hat{\rho}_{MM} = 0,112$  vs. 0,000).

## Matriz de transición laboral — Hombres

Tabla 22: Estimación de las matrices de transición laboral para los hombres bajo el modelo C.  
Los errores estándar se muestran en paréntesis.

$p_{ij}$ (Hombres)	Formal	Informal	Desocupado	Inactivo
Formal	0.791 (0.013)	0.056 (0.004)	0.049 (0.005)	0.102 (0.005)
Informal	0.138 (0.011)	0.403 (0.018)	0.095 (0.008)	0.362 (0.017)
Desocupado	0.146 (0.025)	0.118 (0.017)	0.343 (0.031)	0.392 (0.037)
Inactivo	0.031 (0.004)	0.036 (0.003)	0.030 (0.003)	0.902 (0.021)

## Matriz de transición laboral — Mujeres

Tabla 23: Estimación de las matrices de transición laboral para las mujeres bajo el modelo C.  
Los errores estándar se muestran en paréntesis.

$p_{ij}$ (Mujeres)	Formal	Informal	Desocupado	Inactivo
Formal	0.781 (0.010)	0.039 (0.003)	0.033 (0.003)	0.145 (0.007)
Informal	0.125 (0.012)	0.331 (0.017)	0.073 (0.007)	0.469 (0.018)
Desocupado	0.078 (0.012)	0.086 (0.012)	0.301 (0.028)	0.534 (0.036)
Inactivo	0.017 (0.001)	0.034 (0.002)	0.017 (0.001)	0.931 (0.013)

## Comparación por sexo: Impacto diferencial de la pandemia

Se observan también diferencias significativas en las transiciones hacia la inactividad:

Condición Inicial	Hombres	Mujeres
<b>Formales</b>	10,2 %	14,5 %
<b>Informales</b>	36,2 %	46,9 %
<b>Desocupados</b>	39,2 %	53,4 %
<b>Inactividad persistente</b>	90,2 %	93,1 %

**Conclusión:** Las mujeres sufrieron un mayor impacto en términos de salida del mercado laboral hacia la inactividad.

## Conteos en la muestra

```
base_personas_trim %>% count(pobreza_t1, pobreza_t2, name = "n") %>%
  complete(pobreza_t1 = as.character(1:3),
           pobreza_t2 = as.character(1:3),
           fill = list(n = 0)) %>%
  mutate( pobreza_t1 = case_when(
    pobreza_t1 ==1 ~ "Pobreza extrema",
    pobreza_t1 ==2 ~ "Pobreza no extrema",
    pobreza_t1 ==3 ~ "Fuera de la pobreza"
  ),
  pobreza_t2 = case_when(
    pobreza_t2 ==1 ~ "Pobreza extrema",
    pobreza_t2 ==2 ~ "Pobreza no extrema",
    pobreza_t2 ==3 ~ "Fuera de la pobreza"
  )
) %>%
  pivot_wider(names_from = pobreza_t2, values_from = n, values_fill = 0)
```

## Matriz de observados

pobreza_t1	Pobreza extrema	Pobreza no extrema	Fuera de la pobreza	NA
Pobreza extrema	69	96	0	46
Pobreza no extrema	98	503	796	330
Fuera de la pobreza	0	722	2323	780
NA	17	153	353	121

## Estimación de la población por cambio de estado

```
design_cali %>% group_by(pobreza_t1, pobreza_t2) %>%
  summarise(tol_pob = survey_total() ) %>%
  mutate(total = paste0(round(tol_pob), " (",round(tol_pob_se), ")"),
         tol_pob = NULL, tol_pob_se = NULL) %>%
  mutate( pobreza_t1 = case_when(
    pobreza_t1 ==1 ~ "Pobreza extrema",
    pobreza_t1 ==2 ~ "Pobreza no extrema",
    pobreza_t1 ==3 ~ "Fuera de la pobreza"
  ),
  pobreza_t2 = case_when(
    pobreza_t2 ==1 ~ "Pobreza extrema",
    pobreza_t2 ==2 ~ "Pobreza no extrema",
    pobreza_t2 ==3 ~ "Fuera de la pobreza"
  )
) %>% ungroup() %>%
  pivot_wider(names_from = pobreza_t2,
             values_from = total, values_fill = "0 (0)")
```

## Matriz de estimada

pobreza_t1	Pobreza extrema	Pobreza no extrema	NA	Fuera de la pobreza
Pobreza extrema	113817 (17914)	190139 (28969)	67316 (10740)	0 (0)
Pobreza no extrema	189891 (27999)	930134 (66262)	643658 (41674)	1454773 (59833)
Fuera de la pobreza	0 (0)	1314765 (34368)	1534606 (71370)	4377587 (108473)
NA	31834 (9363)	307645 (24032)	218979 (25131)	665563 (44829)

## Libreria surf

```
# remotes::install_github("guilhermejacob/surf")
library(surf)
design_cali <-
  design_cali %>% mutate(
    pobreza_t1 = factor(
      pobreza_t1,
      levels = c(1, 2, 3),
      labels = c("Pobreza extrema", "Pobreza no extrema",
                 "Fuera de la pobreza")
    ),
    pobreza_t2 = factor(
      pobreza_t2,
      levels = c(1, 2, 3),
      labels = c("Pobreza extrema", "Pobreza no extrema",
                 "Fuera de la pobreza")
    )
  )
```

## Evaluando el modelo C para la muestra

```
model_C <-svyflow(~ pobreza_t1 + pobreza_t2 , model = "C",  
                    design = design_cali ,as.zero.flows = TRUE)
```

- ▶ La mayoría se encontraba fuera de la pobreza (66.7 %).

```
coef(model_C$eta)
```

Pobreza extrema	Pobreza no extrema	Fuera de la pobreza
0.03431872	0.29764073	0.66804055

- ▶ Se observa un leve aumento de la pobreza no extrema y extrema.

```
coef(model_C$gamma)
```

Pobreza extrema	Pobreza no extrema	Fuera de la pobreza
0.03437942	0.28522125	0.68039933

## Resultados del modelo C $p_{ij}$

- ▶ El 38.3 % de los pobres extremos siguen siéndolo.
- ▶ Movilidad moderada desde pobreza no extrema (6.7 % caen en pobreza extrema).

```
coef(model_C$pij)
```

	Pobreza extrema	Pobreza no extrema	Fuera de la pobreza
Pobreza extrema	0.3713847	0.6286153	0.0000000
Pobreza no extrema	0.0726848	0.3640390	0.5632762
Fuera de la pobreza	0.0000000	0.2324637	0.7675363

## Evaluando el modelo C para la muestra de hombres

```
model_C_H <- svyflow(~ pobreza_t1 + pobreza_t2 , model = "C",
                      design = design_cali  %>% filter(sexo == "Hombre") ,
                      as.zero.flows = TRUE)
```

- En el segundo periodo (T2), la mayoría de los hombres se ubican fuera de la pobreza (67.0 %), mientras que una proporción significativamente menor se encuentra en situación de pobreza no extrema (29.7 %) y una proporción aún más baja en pobreza extrema (3.3 %).

Pobreza extrema	Pobreza no extrema	Fuera de la pobreza
0.03088843	0.29587518	0.67323639

- En contraste, en el primer periodo (T1), se observa una mayor proporción de hombres en situación de pobreza no extrema (26.7 %) en comparación con la pobreza extrema (2.9 %).

Pobreza extrema	Pobreza no extrema	Fuera de la pobreza
0.03135252	0.26854632	0.70010116

## Resultados del modelo C para hombres $p_{ij}$

- ▶ Los hombres en pobreza extrema tienen una probabilidad del 78.84 % de permanecer allí.
- ▶ Los hombres en pobreza no extrema presentan una cierta movilidad hacia la pobreza extrema (5.55 %), aunque la mayoría (35.15 %) se mantiene en su estado.

```
coef(model_C_H$pij)
```

	Pobreza extrema	Pobreza no extrema	Fuera de la pobreza
Pobreza extrema	0.3815049	0.6184951	0.0000000
Pobreza no extrema	0.0661375	0.3722846	0.5615779
Fuera de la pobreza	0.0000000	0.2068994	0.7931006

## Evaluando el modelo C para la muestra de mujeres

```
model_C_M <- svyflow(~ pobreza_t1 + pobreza_t2 , model = "C",
                      design = design_cali  %>% filter(sexo == "Mujer") ,
                      as.zero.flows = TRUE)
```

- ▶ La gran mayoría de las mujeres no se encuentran en situación de pobreza al inicio, aunque cerca del 33.51 % se encuentra en alguna forma de pobreza (extrema o no extrema).

Pobreza extrema	Pobreza no extrema	Fuera de la pobreza
0.03747634	0.29953639	0.66298727

- ▶ Se observa un aumento en la pobreza entre las mujeres

Pobreza extrema	Pobreza no extrema	Fuera de la pobreza
0.03733714	0.30025115	0.66241171

## Resultados del modelo C para mujeres $p_{ij}$

- ▶ Alta persistencia de la pobreza extrema entre mujeres: 39.69 % permanecen en ese estado.
- ▶ Las mujeres fuera de la pobreza tienen una probabilidad del 25.64 % de caer en pobreza no extrema.

```
coef(model_C_M$pij)
```

	Pobreza extrema	Pobreza no extrema	Fuera de la pobreza
Pobreza extrema	0.3689701	0.6310299	0.0000000
Pobreza no extrema	0.0784862	0.3578122	0.5637016
Fuera de la pobreza	0.0000000	0.2555474	0.7444526

¡Gracias!

*Email:* andres.gutierrez@cepal.org

## Referencias

- Feinberg, Stephen, y Elizabeth Stasny. 1983. «Estimating monthly gross flows in labour force participation». *Survey Methodology* 9 (1): 77-102.
- Gutiérrez, H. A. 2014. «The estimation of gross flows in complex surveys with random nonresponse». *Survey Methodology* 40 (2): 285-321.
- Jacob, Guilherme. 2020. *surf: Survey-based Gross-Flow Estimation*.
- LaRoche, Silvia. 2003. *Longitudinal and Cross-Sectional Weighting of the Survey of Labour and Income Dynamics*. Statistics Canada.
- Lynn, P. 2009. *Methodology of longitudinal surveys*. Wiley series en survey methodology. Wiley.
- Naud, Jean-Francois. 2002. *Combined-panel longitudinal weighting - Survey of Labour and Income Dynamics*. Statistics Canada.
- Verma, Vijay, Gianni Betti, y Giulio Ghellini. 2006. «Cross-sectional and longitudinal weighting in a rotational household panel: applications to EU-SILC», 36.