

Indicadores longitudinales: características y particularidades

Taller de trabajo sobre análisis de datos longitudinales en encuestas complejas

Andrés Gutiérrez, Ph.D.

Septiembre de 2018 - Santiago de Chile

Encuestas en el tiempo

Introducción

- ▶ Las encuestas realizadas de manera repetida en el tiempo son un instrumento ampliamente usado en la investigación social, económica, médica, etc.
- ▶ Dado que la población es dinámica en el tiempo, la planeación y análisis de este tipo de encuestas es desafiante.

Introducción

- ▶ Si la composición de la población y las características de los elementos se considerara fija, una encuesta transversal sería suficiente para realizar estimaciones.
- ▶ Los cambios naturales en las características de la población hacen que realizar una medición simple en un punto del tiempo sea inviable.
- ▶ A pesar de que crezca la dificultad en el diseño, es posible obtener a información más oportuna para la toma de decisiones

Composición de las poblaciones

- ▶ El tiempo hace que la estructura de las poblaciones cambie, sin importar si la constituyen individuos, hogares, familias, negocios, etc.
 - ▶ Las unidades deben ser consideradas como parte de la población de interés cuando nacen, inmigran, alcanzan un umbral predefinido de edad.
 - ▶ Las unidades salen de la población de interés cuando mueren, emigran, o se involucran en instituciones (como el servicio militar).

Composición de las poblaciones

- ▶ Si las unidades son los hogares ocurre algo similar:
 - ▶ Se crean nuevos hogares cuando los jóvenes dejan a sus padres y forman nuevos hogares independientes, o cuando una separación o divorcio escinde una familia en dos.
 - ▶ Los hogares dejan de ser parte de la población objetivo si todos los miembros del hogar mueren o cuando dos hogares se unen a través de un matrimonio o unión libre.

Encuestas transversales

- ▶ Son diseñadas para recolectar información en un punto específico del tiempo, o sobre un periodo de referencia.
 - ▶ Proveen toda la información pertinente acerca de la población en esa situación (tiempo y periodo).
 - ▶ Como resultado, no es posible estimar cambios de ningún tipo, a no ser que se realicen indagaciones retrospectivas.

Esquema

| Muestra | Tiempo | | | | | |
|---------|----------|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | X | | | | | |
| 2 | X | | | | | |
| 3 | X | | | | | |
| 4 | | | | | | |
| 5 | | | | | | |
| 6 | | | | | | |

Figura1: *Diseño en el tiempo de un encuesta transversal*

Encuestas repetidas

- ▶ Estas encuestas se repiten diferentes puntos del tiempo.
 - ▶ Proveen información acerca de la dinámica de la composición de la población en el tiempo.
 - ▶ En cada levantamiento se selecciona una muestra de la población en un tiempo determinado.
 - ▶ No es posible medir cambios individuales.

Esquema

| Muestra | Tiempo | | | | | |
|---------|----------|----------|----------|----------|----------|----------|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | X | | | | | |
| 2 | | X | | | | |
| 3 | | | X | | | |
| 4 | | | | X | | |
| 5 | | | | | X | |
| 6 | | | | | | X |

Figura2: *Diseño en el tiempo de un encuesta repetida*

Encuestas panel

- ▶ Son diseñadas para recolectar información periódica sobre la misma muestra en diferentes puntos del tiempo.
 - ▶ Las unidades de muestreo son las mismas en los diferentes periodos de tiempo y se miden las mismas variables en cada levantamiento.
 - ▶ Este tipo de encuestas permite medir cambios individuales y cambios netos sobre la población de la que se seleccionó la muestra.
 - ▶ Este tipo de encuestas no capta los cambios de estructura en la población actual.

Esquema

| Muestra | Tiempo | | | | | |
|---------|----------|----------|----------|----------|----------|----------|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | X | X | X | X | X | X |
| 2 | | | | | | |
| 3 | | | | | | |
| 4 | | | | | | |
| 5 | | | | | | |
| 6 | | | | | | |

Figura3: *Diseño en el tiempo de un encuesta de panel*

Encuestas de panel dividido

- ▶ Son una combinación del panel puro y del diseño repetido y su objetivo es realizar inferencia acerca de los cambios de una cohorte a través del tiempo y del cambio en estructura de la población actual.
 - ▶ Se realiza el seguimiento continuo, periódico y sistemático de una muestra, y a la vez, cada levantamiento incluye nuevos elementos seleccionados de la población actual.
 - ▶ Cubre con eficiencia la mayoría de indicadores de interés en un levantamiento longitudinal.

Esquema

| Muestra | Tiempo | | | | | |
|---------|----------|----------|----------|----------|----------|----------|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | X | X | X | X | X | X |
| 2 | X | | | | | |
| 3 | | X | | | | |
| 4 | | | X | | | |
| 5 | | | | X | | |
| 6 | | | | | X | |
| 7 | | | | | | X |

Figura4: *Diseño en el tiempo de un encuesta de panel dividido*

Encuestas de panel rotativo

- ▶ Estas encuestas incorporan nuevos elementos de la población y a la vez mantienen elementos comunes con mediciones anteriores.
 - ▶ En las encuestas panel hay traslape completo entre las muestras de dos puntos en el tiempo. En las encuestas rotativas existe un traslape parcial.
 - ▶ Se reduce el efecto del panel (sobre la población inicial) y el efecto de la pérdida de muestra.
 - ▶ La inclusión de nuevos elementos en la muestra provee información pertinente del cambio en la composición de la población.

Esquema

| Muestra | Tiempo | | | | | |
|---------|----------|----------|----------|----------|----------|----------|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | X | | | | X | X |
| 2 | X | X | | | | X |
| 3 | X | X | X | | | |
| 4 | | X | X | X | | |
| 5 | | | X | X | X | |
| 6 | | | | X | X | X |

Figura5: *Diseño en el tiempo de un encuesta de panel rotativo*

Labor Force Survey (Canadá)

| Mes | Turno de Rotación 6 grupos se entrevistan cada mes | | | | | | | | | |
|-----|---|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | G | H | I | J |
| 1 | 1 | | | | | | | | | |
| 2 | 2 | 1 | | | | | | | | |
| 3 | 3 | 2 | 1 | | | | | | | |
| 4 | 4 | 3 | 2 | 1 | | | | | | |
| 5 | 5 | 4 | 3 | 2 | 1 | | | | | |
| 6 | 6 | 5 | 4 | 3 | 2 | 1 | | | | |
| 7 | | 6 | 5 | 4 | 3 | 2 | 1 | | | |
| 8 | | | 6 | 5 | 4 | 3 | 2 | 1 | | |
| 9 | | | | 6 | 5 | 4 | 3 | 2 | 1 | |
| 10 | | | | | 6 | 5 | 4 | 3 | 2 | 1 |
| 11 | | | | | | 6 | 5 | 4 | 3 | 2 |
| 12 | | | | | | | 6 | 5 | 4 | 3 |
| 13 | | | | | | | | 6 | 5 | 4 |
| 14 | | | | | | | | | 6 | 5 |
| 15 | | | | | | | | | | 6 |
| 16 | | | | | | | | | | |

Figura6: *Esquema de rotación de la muestra (los hogares son observados 6 veces consecutivas)*

Labor Force Survey (Australia)

| Mes | Turno de Rotación | | | | | | | | | |
|-----|----------------------------------|---|---|---|---|---|---|---|---|---|
| | 8 grupos se entrevistan cada mes | | | | | | | | | |
| | A | B | C | D | E | F | G | H | I | J |
| 1 | 1 | | | | | | | | | |
| 2 | 2 | 1 | | | | | | | | |
| 3 | 3 | 2 | 1 | | | | | | | |
| 4 | 4 | 3 | 2 | 1 | | | | | | |
| 5 | 5 | 4 | 3 | 2 | 1 | | | | | |
| 6 | 6 | 5 | 4 | 3 | 2 | 1 | | | | |
| 7 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | | | |
| 8 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | | |
| 9 | | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | |
| 10 | | | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
| 11 | | | | 8 | 7 | 6 | 5 | 4 | 3 | 2 |
| 12 | | | | | 8 | 7 | 6 | 5 | 4 | 3 |
| 13 | | | | | | 8 | 7 | 6 | 5 | 4 |
| 14 | | | | | | | 8 | 7 | 6 | 5 |
| 15 | | | | | | | | 8 | 7 | 6 |
| 16 | | | | | | | | | 8 | 7 |
| 17 | | | | | | | | | | 8 |
| 18 | | | | | | | | | | |

Figura7: Esquema de rotación de la muestra (los hogares son observados 8 veces consecutivas)

Current Population Survey (Estados Unidos)

| Mes | Turno de Rotación | | | | | | | | | | | | | | | | | |
|-----|----------------------------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 4 grupos se entrevistan cada mes | | | | | | | | | | | | | | | | | |
| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R |
| 1 | 1 | | | | | | | | | | | | | | | | | |
| 2 | 2 | 1 | | | | | | | | | | | | | | | | |
| 3 | 3 | 2 | | | | | | | | | | | | | | | | |
| 4 | 4 | 3 | 1 | 1 | | | | | | | | | | | | | | |
| 5 | | 4 | 2 | 2 | 1 | | | | | | | | | | | | | |
| 6 | | | 3 | 3 | 2 | 1 | | | | | | | | | | | | |
| 7 | | | 4 | 4 | 3 | 2 | 1 | | | | | | | | | | | |
| 8 | | | | | 4 | 3 | 2 | 1 | | | | | | | | | | |
| 9 | | | | | | 4 | 3 | 2 | 1 | | | | | | | | | |
| 10 | | | | | | | 4 | 3 | 2 | 1 | | | | | | | | |
| 11 | | | | | | | | 4 | 3 | 2 | 1 | | | | | | | |
| 12 | | | | | | | | | 4 | 3 | 2 | 1 | | | | | | |
| 13 | 5 | | | | | | | | | 4 | 3 | 2 | 1 | | | | | |
| 14 | 6 | 5 | | | | | | | | | 4 | 3 | 2 | 1 | | | | |
| 15 | 7 | 6 | 5 | | | | | | | | | 4 | 3 | 2 | 1 | | | |
| 16 | 8 | 7 | 6 | 5 | | | | | | | | | 4 | 3 | 2 | 1 | | |
| 17 | | 8 | 7 | 6 | 5 | | | | | | | | | 4 | 3 | 2 | 1 | |
| 18 | | | 8 | 7 | 6 | 5 | | | | | | | | | 4 | 3 | 2 | 1 |
| 19 | | | | 8 | 7 | 6 | 5 | | | | | | | | | 4 | 3 | 2 |
| 20 | | | | | 8 | 7 | 6 | 5 | | | | | | | | | 4 | 3 |
| 21 | | | | | | 8 | 7 | 6 | 5 | | | | | | | | | 4 |
| 22 | | | | | | | 8 | 7 | 6 | 5 | | | | | | | | |
| 23 | | | | | | | | 8 | 7 | 6 | 5 | | | | | | | |
| 24 | | | | | | | | | 8 | 7 | 6 | 5 | | | | | | |

Figura8: Esquema de rotación de la muestra (los hogares son observados en una rotación 4-8-4)

Ejemplos de indicadores de interés y su
relación con los tipos de encuestas

Estimación de parámetros poblacionales en distintos puntos del tiempo

- ▶ Ejemplo:
 - ▶ Ingreso per cápita promedio por área (rural - urbano) en las regiones de Chile.
- ▶ Encuestas aptas:
 - ▶ Transversales, repetidas, panel, rotativas y panel dividido.

Estimación de cambios netos

- ▶ Ejemplo:
 - ▶ Diferencia en el número de ocupados de la fuerza de trabajo entre el segundo trimestre de 2018 y el primer trimestre de 2018 en Chile.
- ▶ Encuestas aptas:
 - ▶ Repetidas, panel, rotativas y panel dividido.
- ▶ Encuestas no aptas:
 - ▶ Transversales.

Estimación de cambios brutos y componentes individuales

- ▶ Ejemplo:
 - ▶ Porcentaje de personas ocupadas en el segundo trimestre de 2018 que estuvieron desocupadas en el primer trimestre de 2018 en la Región Metropolitana de Chile.
- ▶ Encuestas aptas:
 - ▶ Panel, rotativas y panel dividido.
- ▶ Encuestas no aptas:
 - ▶ Transversales, repetidas.

Estimación de tendencias mediante la acumulación de datos

- ▶ Ejemplo:
 - ▶ Tendencia del ingreso medio por cohorte de egresados de un programa de educación universitaria.
- ▶ Encuestas aptas:
 - ▶ Panel y panel dividido.
- ▶ Encuestas no aptas:
 - ▶ Transversales, repetidas y rotativas.

Incidencia de eventos en un periodo de tiempo

- ▶ Ejemplo:
 - ▶ Proporción de mujeres que fueron víctimas de un evento de violencia en los últimos seis meses.
- ▶ Encuestas aptas:
 - ▶ Transversales, repetidas, panel, rotativas y panel dividido.

Incidencia de eventos raros en el tiempo

- ▶ Ejemplo:
 - ▶ Estimación de factores asociados con el complejo teniosis-cisticercosis en las personas de la región rural de Chile.
- ▶ Encuestas aptas:
 - ▶ Repetidas, rotativas y panel dividido.
- ▶ Encuestas no aptas:
 - ▶ Transversales y panel.

Relación entre variables

- ▶ Ejemplo:
 - ▶ Correlación entre las tasas de crimen y de arresto efectivo en las comunas de la Región Metropolitana en Chile.
- ▶ Encuestas aptas:
 - ▶ Transversales, repetidas, panel, rotativas y panel dividido.

Efecto del tipo de encuesta en la eficiencia de
los indicadores

Cambios netos (parámetro de interés)

Considere el cambio neto de la media de la variable de interés x en dos periodos de tiempo (t_2 y t_1)

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

Cambios netos (estimador de muestreo)

El cambio en los dos periodos es estimado de forma aproximadamente insesgada mediante la siguiente expresión:

$$\hat{\Delta} = \hat{x}_2 - \hat{x}_1 = \frac{\sum_{k \in s_2} \frac{x_k}{\pi_k}}{\sum_{k \in s_2} \frac{1}{\pi_k}} - \frac{\sum_{k \in s_1} \frac{x_k}{\pi_k}}{\sum_{k \in s_1} \frac{1}{\pi_k}}$$

En donde s_2 y s_1 representan las muestras seleccionadas en los periodos de interés y π_k es la probabilidad de inclusión del elemento k .

Cambios netos (precisión y eficiencia)

La varianza del estimador de cambio se calcula mediante la siguiente expresión:

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

En general, el último término se puede expresar como

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

En donde T_2 y T_1 representan el porcentaje de muestra común que se traslapa en ambos levantamientos y R_{12} representa la correlación de la variable de interés x en los periodos observados.

Cambios netos (precisión y eficiencia)

Suponiendo que la variación de la variable de interés es homogénea en ambos periodos y que el traslape es común por diseño, entonces la expresión de la varianza se reduce de la siguiente manera:

$$Var(\hat{\Delta}) = 2Var(\hat{x}) - 2\sqrt{Var(\hat{x})}TR = 2Var(\hat{x})(1 - TR)$$

Efecto del tipo de encuesta en la precisión

Nótese que la varianza de este indicador cambiará de acuerdo al tipo de encuesta que se elija:

- ▶ Encuesta repetida: en donde $T = 0$ y

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

.

- ▶ Encuesta de panel, en donde $T = 1$, $R > 0$ y

$$Var(\hat{\Delta}) = 2Var(\hat{x})(1 - R)$$

.

- ▶ Encuesta rotativa: en donde $T \neq 0$, $R > 0$ y

$$Var(\hat{\Delta}) = 2Var(\hat{x})(1 - TR)$$

.

Efecto del tipo de encuesta en el tamaño de muestra

Si se supone que la **correlación es positiva** para la variable en los dos periodos de tiempo, entonces se tiene la siguientes conclusión:

$$2Var(\hat{\bar{x}})(1 - R) < 2Var(\hat{\bar{x}})(1 - TR) < 2Var(\hat{\bar{x}})$$

Es decir que se necesita **menos** tamaño de muestra para medir los cambios netos usando un diseño panel que un diseño sin traslape. Un camino medio es el diseño rotativo.

Promedio trimestral (parámetro de interés)

Considere una encuesta continua y mensual en donde se quiere estimar el promedio trimestral de la variable de interés x en tres periodos de tiempo (t_3 , t_2 y t_1)

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

Cambios netos (estimador de muestreo)

Un estimador del promedio trimestral que es estimado de forma aproximadamente insesgada mediante la siguiente expresión:

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

En donde s_3 , s_2 y s_1 representan las muestras seleccionadas en los periodos de interés y π_k es la probabilidad de inclusión del elemento k .

Promedio trimestral (precisión y eficiencia)

La varianza del estimador del promedio trimestral se calcula mediante la siguiente expresión:

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

Cambios netos (precisión y eficiencia)

Suponiendo que la variación de la variable de interés es homogénea en ambos periodos y que el traslape es común por diseño y que los errores de muestreo son débilmente estacionarios entre dos y tres meses, entonces la expresión de la varianza se reduce de la siguiente manera:

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

En donde R es la correlación de la variable de interés en dos y tres meses (asumida homogénea).

Efecto del tipo de encuesta en la precisión

Nótese que la varianza de este indicador cambiará de acuerdo al tipo de encuesta que se elija:

- ▶ Encuesta repetida: en donde $T = 0$ y

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

- ▶ Encuesta de panel, en donde $T = 1$, $R > 0$ y

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

- ▶ Encuesta rotativa: en donde $T \neq 0$, $R > 0$ y

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

Efecto del tipo de encuesta en el tamaño de muestra

Si se supone que la **correlación es positiva** para la variable en los tres periodos de tiempo, entonces se tiene la siguiente conclusión:

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

Es decir que se necesita **más** tamaño de muestra para estimar un promedio trimestral usando un diseño panel que un diseño sin traslape.

Estimando indicadores actuales usando
encuestas actuales y pasadas

Estimando en un punto del tiempo

- ▶ Existen enfoques de estimación que permiten apoyarse en los datos de encuestas anteriores para obtener un estimador del indicador actual (medias, proporciones, etc.).
 - ▶ Se crea un procedimiento iterativo que actualiza la estimación de interés a medida que se realiza un nuevo levantamiento en la encuesta.
 - ▶ Como insumo fundamental es necesario estimar el cambio que ocurre entre los periodos consecutivos en la parte traslapada y no traslapada de las muestras.

El estimador compuesto (K)

Este enfoque utiliza un promedio ponderado entre la estimación puntual, calculada con los datos actuales, y una estimación indirecta, usando los datos pasados y la estimación del cambio sobre la muestra traslapada.

$$\textit{Compuesto}^K = \textit{Actual} + (\textit{Pasado} + \textit{Cambio}^M)$$

La forma de este estimador es:

$$\hat{t}_2^K = K\hat{t}_2 + (1 - K)(\hat{t}_1 + (\hat{t}_2^M - \hat{t}_1^M))$$

El estimador compuesto (AK)

Este estimador parte del estimador K y añade un término de error relacionado con la diferencia de la estimación actual de las partes no traslapadas y traslapadas.

$$Compuesto^{AK} = Actual + (Pasado + Cambio^M) + (Actual^U - Actual^M)$$

La forma de este estimador es:

$$\hat{t}_2^{AK} = K\hat{t}_2 + (1 - K)(\hat{t}_1 + (\hat{t}_2^M - \hat{t}_1^M)) + A(\hat{t}_2^U - \hat{t}_2^M)$$

Estimador compuesto iterativo

Se usa como insumo principal la estimación compuesta del periodo anterior y parte del estimador K y añade un término de error relacionado con la diferencia de la estimación actual de las partes no traslapadas y traslapadas. La forma de este estimador es:

$$\hat{t}_2^{AK} = K\hat{t}_1^{AK} + (1 - K)(\hat{t}_1 + (\hat{t}_2^M - \hat{t}_1^M)) + A(\hat{t}_2^U - \hat{t}_2^M)$$

Consideraciones en la producción de las cifras

A pesar de que los estimadores compuestos utilizan información auxiliar que mejora la eficiencia de las cifras, una pregunta de interés es

¿Cómo estandarizar este procedimiento a la producción usual de las estadísticas oficiales en los INE?

Nótese que los procedimientos estándar están supeditados al principio de representatividad y al uso de los factores de expansión.

Calibración de estimadores compuestos

- ▶ Es posible utilizar los principios de los estimadores de calibración para hacer posible la estimación compuesta iterativa en una base de datos actual con un sólo conjunto de factores de expansión.
 - ▶ Es posible calibrar con la intención de estimar indicadores de nivel (porcentajes, medias o totales) o indicadores de cambio (diferencias de medias, diferencias de porcentajes).
 - ▶ Ambas opciones guía a diferentes conjuntos de factores de expansión.

Estimadores de calibración compuestos (indicadores de nivel)

Si el interés principal son los indicadores de nivel, se recomienda utilizar la siguiente variable de información auxiliar:

$$z_k = \begin{cases} \bar{x}^1, & \text{para todo } k \in s^U \\ x_k^1, & \text{para todo } k \in s^M \end{cases}$$

Sujeto a la siguiente restricción

$$\hat{t}_1 = \sum_{k \in s} w_k z_k$$

Estimadores de calibración compuestos (indicadores de cambio)

Si el interés principal son los indicadores de cambio, se recomienda utilizar la siguiente variable de información auxiliar:

$$z_k = \begin{cases} x_k^2, & \text{para todo } k \in s^U \\ x_k^2 - T^{-1}(x_k^1 - x_k^2), & \text{para todo } k \in s^M \end{cases}$$

Sujeto a la siguiente restricción

$$\hat{t}_1 = \sum_{k \in s} w_k z_k$$

Pruebas de hipótesis sobre indicadores longitudinales

Pruebas sobre cambios netos

Considere el siguiente sistema de hipótesis, sobre la diferencia en dos tiempos (cambio neto) consecutivos.

$$H_0 : \theta_2 - \theta_1 = 0 \quad vs \quad H_1 : \theta_2 - \theta_1 \neq 0$$

Existen varias consideraciones que deben realizarse sobre los subgrupos en los que se intenta realizar la inferencia estadística.

Estadístico de prueba

Para llevar a cabo la prueba de hipótesis trabajamos con el siguiente estimador de diferencias:

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

La varianza asociada a este estimador está dada por

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

El término de covarianza se puede escribir como

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

Principio fundamental

Existen muchos escenarios de comparación que son de interés cuando se analizan datos de una encuesta de empleo. Estas comparaciones se hacen mas complejas cuando se incluye en el análisis el diseño en panel de la encuesta.

A no ser que los dos estimadores puntuales estén compuestos de observaciones provenientes de un conjunto disyunto de UPMs, el término de covarianza no será nulo.

Covarianza en comparaciones mensuales

Suponga que se quiere comparar la tasa de desempleo nacional entre dos meses consecutivos en una encuesta repetida.

- ▶ Existe independencia en el muestreo de los dos meses consecutivos y por lo tanto el porcentaje de traslape de muestra entre los dos meses (que por diseño es nulo) es igual a cero.
- ▶ Por lo tanto, $T_1 = T_2 = 0$. Luego, el término de la covarianza se anula. En resumen, la varianza del estimador en este caso sería igual a:

$$Var(\hat{d}) = Var(\hat{\theta}_2) + Var(\hat{\theta}_1)$$

Covarianza en comparaciones trimestrales o anuales

Suponga que se quiere comparar la tasa de desempleo nacional entre dos meses consecutivos en una encuesta rotativa.

- ▶ En este escenario no existe independencia, puesto que la estructura del panel garantiza un traslape (suponga $T_1 = T_2 = 05$).
- ▶ Existe una correlación natural entre las viviendas comunes en el panel que se midieron en los periodos de interés, por lo tanto $R_{1,2} \neq 0$. Note que esta correlación se calcula sobre los individuos comunes en el panel.
- ▶ En resumen, el término de covarianza en este caso sería igual a:

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

Covarianza en comparaciones de un mismo mes

Suponga que se quiere comparar la tasa de desempleo entre hombres y mujeres en un mismo mes.

- ▶ No existe independencia en el muestreo de hombres y mujeres puesto que estos grupos no son estratos de muestreo. En este caso T_1 es la proporción de hombres en la muestra y T_2 es la proporción de mujeres en la muestra. Nótese que $T_1 \neq T_2$.
- ▶ Existe una correlación natural entre las UPMs que fueron seleccionadas y que contienen tanto a hombres como a mujeres, por lo tanto $R_{12} \neq 0$.
- ▶ En resumen, el término de covarianza en este caso sería igual a:

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

Covarianza en comparaciones de un mismo mes

Suponga que se quiere comparar la tasa de desempleo entre dos ciudades principales en un mismo mes.

- ▶ Existe independencia en el muestreo de las dos ciudades porque la selección es independiente en cada ciudad, puesto que ambas ciudades son estratos de muestreo.
- ▶ En este caso T_1 es la proporción de personas de una ciudad y T_2 es la proporción de personas de la otra ciudad. Nótese que $T_1 \neq T_2$.
- ▶ No existe una correlación entre las UPMs que fueron seleccionadas en las ciudades porque la selección fue independiente, por lo tanto $R_{12} = 0$ y el término de covarianza es nulo y por ende la varianza del estimador sería igual a:

$$Var(\hat{d}) = Var(\hat{\theta}_2) + Var(\hat{\theta}_1)$$

Estadístico de prueba

Una vez se ha definido la estructura de varianza del estimador de interés, el siguiente paso es definir si la siguiente cantidad es distinta de cero para determinar si el parámetro ha cambiado entre grupos o a lo largo del tiempo.

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

Este estadístico de prueba sigue una distribución *t-student* con df grados de libertad:

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

Para cerrar...

Temas que merecen ser discutidos

Es muy amplio el contenido del diseño y análisis de las encuestas longitudinales, pero existen algunos tópicos que considero muy importantes a la hora de abordar una encuesta de este tipo:

1. Ausencia de respuesta: como un reto metodológico y de diseño.
2. Creación de categorías de respuesta y definición del tipo de respuesta longitudinal.
3. Construcción de un sistema de ponderación (factores de expansión) longitudinal en encuestas continuas con panel rotativo.
4. Calibración longitudinal de factores de expansión longitudinales.
5. Estimación de errores de muestreo para indicadores longitudinales en desagregaciones de interés.

¡Gracias!

Andrés Gutiérrez

Experto Regional en Estadísticas Sociales

Comisión Económica para América Latina y el Caribe

ONU-CEPAL

Email: andres.gutierrez@cepal.org