CEPAL OIT INEC Ecuador

Taller Regional

Las Encuestas de Hogares en América Latina: Estado de situación y prospectiva

Diseños de muestra para encuestas longitudinales

Fernando Medina

CEPAL

14 de octubre, 2015

Objetivos de las encuestas longitudinales

- El ***objetivo central*** de los estudios longitudinales es generar información que permita conocer la evolución dinámica de fenómenos económicos y sociales.

- Elegir el mejor diseño estadístico requiere ***evaluar los costos y beneficios*** de las distintas opciones, dependiendo de los objetivos de la investigación que se desea realizar.

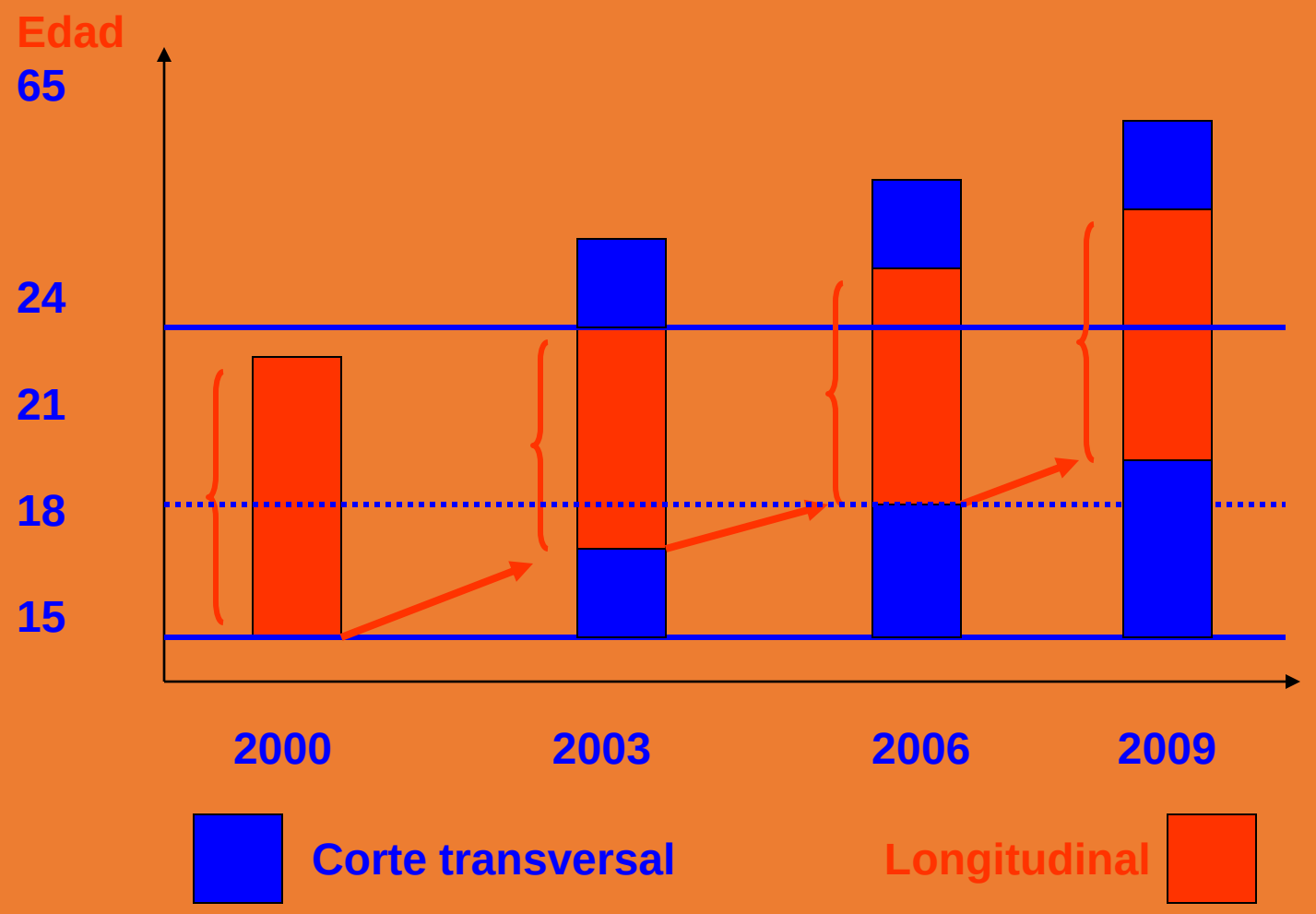
- El principal atributo de las investigaciones longitudinales es generar ***información repetida*** para un conjunto de observaciones de interés.

Objetivos

- Una ***encuesta longitudinal*** tiene como propósito recolectar datos sobre una ***misma muestra en distintos momentos*** del tiempo.

- El objetivo de este tipo de investigaciones es hacer el ***seguimiento*** de un grupo de observaciones ***para medir diferencias*** intertemporales.

Diferencia entre una encuesta de corte transversal y una longitudinal



Encuestas y datos longitudinales

- Las ***encuestas longitudinales*** son la fuente que genera los datos longitudinales.

- No obstante, también se pueden generar ***datos longitudinales*** sin necesidad de llevar a cabo estudios longitudinales.

- Esto se logra a partir de ***investigaciones*** de carácter ***retrospectivo***, que indagan acerca de variables de interés en el pasado. Ejemplo, los estudios sobre fecundidad y mortalidad.

Ventajas de los datos longitudinales

- Las principales ventajas de los datos longitudinales son de carácter analítico y se refieren a los siguientes aspectos:

✓ Análisis del cambio bruto

✓ Análisis del cambio medio por unidad de observación

✓ Análisis de la estabilidad del fenómeno de estudio

✓ Análisis del tiempo de ocurrencia de un suceso: Persistencia

✓ Análisis de la naturaleza ordinal de los sucesos para identificar relaciones de causalidad

Análisis del cambio bruto

- El análisis del cambio bruto es uno de los intereses centrales en el uso de datos longitudinales.
- Las encuestas de corte transversal se utilizan frecuentemente para estimar el cambio neto, por ejemplo, en la tasa de desempleo o en la incidencia de la pobreza.
- No obstante, sólo a partir de una encuesta longitudinal es posible explicar hasta que punto el cambio neto, se explica por los distintos elementos del cambio bruto.
- En el tema del desempleo, por ejemplo, los estudios longitudinales ayuda a conocer si a partir de dos tasas de desempleo similares las personas ubicadas en condición de desempleo abierto son las mismas en ambos periodos, o existen flujos opuestos o equivalentes dentro o fuera del desempleo

Análisis del cambio individual

- El análisis del cambio individual es relevante, independientemente del cambio neto.
- Este tipo de análisis es relevante para la evaluación de impacto de políticas públicas en favor del empleo, el mejoramiento de los ingresos individuales y familiares, así como para monitorear la evolución de los indicadores de pobreza.
- El estudio de los cambios individuales sólo se puede llevar a cabo a partir de información longitudinal.
- El análisis de datos longitudinales en este contexto, además de qie permite estimar el cambio neto, da la posibilidad de identificar los factores determinantes que explican los cambios observados en las unidades de observación.

Duración de los fenómenos

- En el contexto del diseño y la evaluación de las políticas públicas es relevante conocer la duración de los episodios o fenómenos que se analizan.

- En el ámbito del mercado laboral, por ejemplo, la duración de los episodios de desocupación son particularmente importantes para entender la dinámica del mercado de trabajo y su relación con el ciclo económico, así como para la generación de ingresos y sus vínculos con la trayectoria asumida por los indicadores de pobreza.

- Desde la óptica analítica, los modelos de efectos aleatorios y el análisis de supervivencia son las técnicas estadísticas y econométricas que se aplican para estudiar la propensión al cambio y los factores que lo determinan.

Tipos de investigaciones longitudinales

- Retrospectivas

- Panel
- ✓ Cohorte de hogares
- ✓ Cohorte de personas
- ✓ Paneles con rotación

- Registros administrativos

Tipología de los diseños de longitudinales

- En una investigación de panel la unidad de observación es seguida en el tiempo y la información se recolecta en forma secuencial en distintos periodos.

- **Paneles fijos o indefinidos**: Se refiere a investigaciones que recolectan información sobre una misma unidad de observación en múltiples ocasiones.

- Toda vez que se selecciona la muestra no se permite la inclusión de nuevas unidades de observación. Se refiere a una **población cerrada que sólo admite salidas ("muerte" o migración) pero no entradas.**

- **En este tipo de diseños se puede o no seguir en el tiempo y el espacio a las unidades de observación.**

Tipología de los diseños de longitudinales (cont.)

- **Paneles con duración fija con nuevos ingresos**. Asume la misma lógica del diseño de paneles fijos, con la diferencia que en este caso si se admiten ingresos.

- En el caso de las encuestas en hogares, por ejemplo, en el panel se incluyen unidades de observación (personas) que son nuevos residentes en el hogar.

- Uno de las ventajas de este tipo de diseños es que al incorporar nuevas unidades de observación es posible mejorar las estimaciones de población de corte transversal.

- Asimismo, permite reconocer a las unidades de observación longitudinal que permanecen en el panel original.

Tipología de los diseños de longitudinales (cont.)

- **Paneles repetidos**: Bajo este diseño se dispone de un conjunto de muestras que pueden o no traslaparse en el tiempo.

- Cada panel (submuestra) tiene el propósito explícito de representar de manera adecuada a una población de referencia en un momento determinado.

- Cada panel consiste en una muestra que representa a una ***cohorte con características particulares***.

Tipología de los diseños de longitudinales (cont.)

- **Paneles con rotación**: La lógica que sustenta este diseño asume que deben sustituirse proporciones de las unidades de observación en el tiempo.
- Habitualmente, bajo este diseño cada unidad de observación permanece en la muestra el mismo número de periodos.
- Las diversas muestras permiten generar estimaciones de corte transversal y longitudinal.
- Lo anterior, debido a que los paneles de rotación son equilibrados y en cualquier momento del tiempo representan a la misma población de referencia.
- Este tipo de diseños se utilizan cuando los objetivos se asocian al interés de generar estimaciones transversales y longitudinales.

Tipología de los diseños de longitudinales (cont.)

- En la medida de que el porcentaje de traslape sea mayor, esto contribuye a la reducción de la varianza en los estimadores de cambio.

- En el caso de algunos de los países de la región, la proporción de traslape en los distintos turnos de rotación de la muestra (20 0 25%), garantiza un porcentaje importante de solapamiento que contribuye a la reducción de la varianza en los estimadores de cambio.

Tipología de los diseños de longitudinales (cont.)

- **Paneles divididos (combinados)**: Este diseño combina muestras transversales y longitudinales en cada periodo.

- Bajo esta modalidad el diseño más común complementa una encuesta longitudinal con una muestra de corte transversal en un omento determinado.

- Este tipo de diseños se aplican con el propósito de mejorar la precisión de las estimaciones de corte transversal.

Población de referencia

- En los estudios longitudinales la ***identificación de la población objetivo*** es de gran importancia para la construcción de los ponderadores.

- Por ejemplo, en una investigación con dos o más mediciones en donde se examina la dinámica de un fenómeno de interés entre t_1 y t_2 la población de referencia puede ser la muestra seleccionada en t_1 o la población que se observó en t_2.

- Si sólo interesa estudiar a los seleccionados en t_1 estamos en presencia de una población que **admite salidas pero no ingresos**. Por consiguiente, el tamaño de muestra en t_1 será menor o igual al seleccionado en t_2 : $n_2 \leq n_1$

Población de referencia (cont.)

- Otra opción que se presenta es que la población de referencia se defina como la **intersección entre la población de corte longitudinal y transversal.**

- Bajo esta lógica no se aceptan ni entradas ni salidas de las unidades de observación. Observe que ante esta variante se pueden presentar problemas en los estimadores y es probable que se tenga que recurrir a técnicas de calibración para generar estimadores insesgados y controlar su varianza.

- Observe que la intersección de las unidades de observación se modifica dependiendo de los periodos que se desean comparar.

Población de referencia (cont.)

- Otra opción es definir a la población de referencia como la **unión entre la población de corte longitudinal y transversal.**

- Este enfoque se considera el más adecuado para el estudio dinámica de una población, en la medida de que se aceptan ingresos de nuevas unidades de observación con el propósito de evitar sesgos de estimación.

Construcción de los ponderadores en las encuestas longitudinales

- Se reconocen al menos las siguientes razones para ponderar los resultados de una muestra:

✓ Corregir las diferencias en las probabilidades de selección de las unidades muestrales.

✓ Corregir las diferencias en la actualización del marco de muestreo.

✓ Corregir las diferencias en las tasas de cobertura.

✓ Corregir las diferencias en las tasas de respuesta.

✓ Para corregir los efectos de la estratificación y la conglomeración de la muestra.

✓ Para corregir las diferencias entre las estimaciones de la muestra y las proyecciones de población.

Objetivo del uso de los ponderadores

- El objetivo fundamental que justifica el uso de lo ponderadores es garantizar que los grupos de unidades elegidos en la muestra estén representados de la misma manera en que se distribuyen en la población de referencia.

- Las características de los ponderadores en una encuesta longitudinal guardan relación con la población objeto de estudio y con el tratamiento de los patrones de no respuesta observados.

Ponderadores en las encuestas longitudinales

- En muchas de las investigaciones longitudinales se generan dos tipos de ponderadores:

✓ *para estimaciones transversales*

✓ *para generar estimaciones longitudinales*

- Las diferencias entre los ponderadores se genera a partir de la población que se desea representar, lo cual está asociado a los objetivos de la investigación.

Ponderadores en las encuestas longitudinales (cont.)

- Para generar estimaciones entre dos puntos t1 y t2 se presentan las siguientes opciones:

✓ Estimaciones longitudinales con información de los puntos t1 y t2.
✓ Estimadores transversales con los datos de t1.
✓ Estimadores transversales con los datos de t2.

- En todos los casos los estimadores son diferentes con relación a las unidades de observación como a la población de referencia.

Estimadores de interés

- Habitualmente en los diseños longitudinales los estimadores de interés están asociados a los cambios en las variables de estudio.

- En las investigaciones del mercado de trabajo se tienen los siguientes estimadores (entre muchos otros):

✓ *Estimador de cambio en la tasa de desocupación.*

✓ *Estimador de cambio en el volumen de ocupados y desocupados.*

✓ *Estimador de cambio en la duración promedio del desempleo.*

✓ *Estimador de cambio en el ingreso medio de los ocupados.*

Tamaño de muestra

- En el caso de los estudios longitudinales, además de requerir una estimación de la varianza de la variable de interés, es necesario anticipar la pérdida de muestra que se tendrá por la no respuesta en la primera visita y la que se generará debido al desgaste de la muestra en el tiempo.

- Investigaciones efectuadas en países europeos sugieren que la pérdida de muestra en los estudios longitudinales se presenta con mayor frecuencia en la primera y segunda ronda.

- Una situación similar se reporta en el caso de algunosm países de América Latina.

- Conforme a lo anterior, la expresión que se utilice para determinar el tamaño de la muestra debe tener en cuenta la pérdida de muestra por no respuesta y por desgaste del panel (attrition).