UNIVERSIDAD MAYOR DE SAN ANDRÉS
FACULTAD DE CIENCIAS PURAS Y NATURALES

CARRERA DE ESTADÍSTICA

TESIS DE GRADO

DISEÑO DE ENCUESTAS TIPO PANEL ROTATORIO CON TRASLAPE PARCIAL

Tesista: Mercy Crespo Chuquimia
Tutor: Lic. Fernando Rivero Suguiura

La Paz – Bolivia
2011

AGRADECIMIENTOS

En primer lugar un agradecimiento muy especial a mi señor padre Humberto Crespo, por su esfuerzo para darme todo lo necesario, y poder hacer frente a esta vida llena de dificultades.

A mi mamá Hulda, quien además de darme la vida, me brindó su apoyo esencial en todo momento y en especial durante la elaboración de esta tesis. También un agradecimiento especial a mi mamá Geno y a mis hermanas por su aliento y apoyo incondicional.

Mis sinceros agradecimientos al Lic. Fernando Rivero por sus orientaciones y valiosos consejos, así mismo quiero agradecer al Lic. Jaime Pinto y Lic. Nilda Flores por sus aportes en la finalización de este trabajo

A los docentes que me han acompañado durante el largo camino, brindándome siempre su orientación con profesionalismo ético en la adquisición de conocimientos y afianzando mi formación como estudiante universitaria.

A todos ellos …gracias.

A la razón de mi vida: *Lunita*

A mi compañero: *Clifford*

Índice General

Capítulo 1

Introducción

1.1 Antecedentes

Cuando se desea conocer los valores de los parámetros de una población y no se cuenta con los recursos necesarios para realizar un censo entonces se recurre a las encuestas por muestreo, de modo que mediante estas se pueda tener una estimación del valor del parámetro para un tiempo de referencia.

Sin embargo, los parámetros poblacionales pueden ir variando conforme transcurre el tiempo, de tal forma que una encuesta transversal o puntual, solamente genera la estimación del parámetro para un tiempo dado, desconociendo los valores del parámetro antes y después de la encuesta.

Entonces, si el objetivo de una investigación es conocer el comportamiento que tiene un parámetro poblacional a lo largo del tiempo, recurrimos a las encuestas longitudinales, las cuales tienen la característica de reunir datos, a partir de los mismos elementos muestrales, en múltiples ocasiones a través del tiempo.

Las encuestas longitudinales son utilizadas por muchos países alrededor del mundo, con la finalidad de tener un monitoreo constante de lo que ocurre con un valor poblacional de interés. Las encuestas de fuerza de trabajo por ejemplo, son estudios que tienen por objetivo dar estimaciones de la situación de empleo, indicadores de cambio, dinámica de empleo, etc., los cuales sólo se logran a través de este tipo de encuestas.

Podemos citar algunas de las encuestas, por ejemplo: Encuesta Permanente de Hogares EPH llevada a cabo por el INDEC de Argentina (Instituto Nacional de Estadística y Censos), la ENOE que es la Encuesta Nacional de Ocupación y Empleo llevada a cabo por el INEGI (Instituto Nacional de Estadística, Geografía e Informática) de México, y una de las encuestas más antiguas que se ha estado realizando es British Household Panel Survey llevada a cabo en el Instituto de Investigaciones Sociales y Económicas de la Universidad de Essex , con una muestra de los hogares británicos los cuales se entrevistaron por primera vez en 1991. Los miembros de estos hogares originales desde entonces se han seguido y se entrevistó con los años.

Todas ellas son encuestas longitudinales de tipo panel, que permiten estimar cambios a nivel individual (cambio bruto), como también cambios en términos del total de la población (cambio neto), cuyos resultados generan un apoyo para la creación de nuevas políticas y decisiones que se tomarán a partir de ellos.

Sin embargo, una de las principales y más serias limitaciones de este tipo de encuestas, es sin duda el cansancio de los respondientes o pérdida de observaciones a medida que pasa el tiempo, que da como resultado un incremento significativo de la tasa de no respuesta. Las encuestas longitudinales tipo panel con rotación y traslape parcial son diseñadas para seguir a un grupo particular de unidades muestrales por un específico periodo de tiempo, y al mismo tiempo se incluyen nuevas unidades muestrales en orden que van saliendo otras. El hecho de rotar la muestra implica que las personas pertenecientes al panel salgan de la muestra una vez que hayan cumplido el ciclo de permanencia en la encuesta, y no sigan en ella por tiempo indefinido, reduciendo así el desgaste de las unidades seleccionadas.

1.2 Identificación y planteamiento del problema

Al buscar información acerca de las encuestas panel, se pudo identificar que existe muy poca información en lo que respecta a las bondades de este tipo de encuesta y una metodología que de directrices para una planificación y desarrollo de las encuestas tipo panel muy poco utilizadas en nuestro país.

Al no tener disponible esta información se puede incurrir en errores al momento de la planificación y ejecución de la encuesta.

Lo anterior abre una posibilidad de investigar sobre este tema y documentar las bondades y las mejoras en los estimadores que se pueden lograr a través del uso de estas encuestas panel.

1.3 Justificación

Toda investigación cuantitativa se realiza a través de encuestas que tienen sus propias características. Una de ellas es la encuesta panel rotatorio con traslape parcial, que es utilizada principalmente en las encuestas de hogares por los institutos de estadística en varios países.

Una encuesta tipo panel rotatorio recolecta información longitudinal, la que se obtiene a partir de encuestas que se realizan periódicamente a un mismo grupo de personas. Este tipo de encuestas es también una herramienta que permite la realización de estudios sociales que no pueden ser inferidos mediante un estudio transversal. Un panel permite que se midan y se analicen por ejemplo variables micro-económicas de una sociedad. También puede controlar los efectos de las variables que no se ven reflejadas en cierto periodo porque no dependen de la variable tiempo.

Veamos el siguiente cuadro de comparación sobre la información que se puede obtener según tipo de encuesta.

Tabla 1.1

	Tipo de Encuesta				
Tipo de Estimación	**Una vez**	**Repetida no traslapada**	**Repetida traslapada parcialmente**	**Longitudinal sin rotación**	**Longitudinal con rotación**
Para un punto en el tiempo	x	x	x	x	x
Duraciones, transiciones, frecuencia de ocurrencia	x	x	x	x	x
Relación entre características	x	x	x	x	x
Cambio neto		x	x	x	x
Tendencias		x	x	x	x
Eventos raros - datos acumulados		x	x		x
Cambio bruto			x	x	x
Características para largos periodos de tiempo basados en datos acumulados				x	x

Fuente: Panel survey (1989), página 3

1.4 Objetivos

1.4.1 Objetivo General

Desarrollar la metodología para el diseño panel de las encuestas panel rotatorio con traslape parcial considerando aspectos teóricos y prácticos de este tipo de encuestas poco desarrolladas y utilizadas en nuestro país. Posteriormente realizar una aplicación de lo desarrollado en el documento, de modo que el presente trabajo sea útil para investigaciones posteriores que tengan la necesidad de realizar un estudio a través de encuestas con similares características.

1.4.2 Objetivos Específicos

- Mostrar diferentes modelos de rotación para las encuestas panel
- Determinar los principales problemas en estudios por encuestas panel y plantear posibles soluciones
- Analizar el traslape de la muestra en un diseño panel rotatorio
- Analizar los estimadores de cambio y promedio
- Analizar el estimador combinado mejorado con información anterior

1.5 Alcance

El presente trabajo desarrolla de manera general y específica la metodología a seguir en el diseño panel de las encuestas panel rotatorio con traslape parcial. El diseño muestral de una encuesta involucra muchas otros aspectos, los cuales solamente se citarán de manera referencial.

Por otra parte se ven tres tipos de estimación: Estimación del cambio, estimación del promedio y el estimador combinado. Por simplicidad, todos ellos se desarrollan bajo el supuesto de un muestreo aleatorio simple

Capítulo 2

Marco Teórico

2.1 Introducción

En este capítulo se desarrolla de forma específica el diseño panel de una encuesta tipo panel rotatorio con traslape parcial. Sin embargo, el diseño de una encuesta cualquiera no solamente implica realizar el diseño panel, sino también determinar los objetivos de la encuesta, ámbito de la encuesta, etc. Si en realidad se desea toda esta información, entonces se deberán recurrir a otras fuentes para profundizar estos puntos, ya que a continuación se mencionarán los pasos a seguir para el diseño de una encuesta solamente de manera referencial.

2.2 Definiciones generales

Antes de ingresar más a fondo en este capítulo, se ve la pertinencia de definir lo siguiente:

Panel

Un panel es un grupo de unidades muestrales a las cuales se hacen mediciones sucesivas por un tiempo o tiempo indefinido.

Rotación de paneles

El hecho de que se tenga una muestra con rotación de paneles, significa que las unidades muestrales no serán objeto de estudio por tiempo indefinido. Al rotar los paneles, una unidad muestral es estudiada por un número determinado de veces. De esta forma, en cada ronda de la encuesta se incluye un porcentaje de muestra totalmente nueva.

Traslape (β)

El traslape hace referencia a las unidades muestrales que se mantienen de una ronda a otra. Por ejemplo, si en una ronda existe 20 unidades muestrales y de esas 20 unidades muestrales se mantienen 5 unidades muestrales para la siguiente ronda, entonces el traslape de esas dos rondas se calcula de la siguiente manera:

$$\beta = \frac{5}{20} = 0.25$$

En términos porcentuales, el traslape sería de un 25%.

De aquí el término "traslape parcial" se refiere a que no todas las unidades muestrales se mantienen para una ronda posterior, sino que se traslapan solo parte de ellas.

Ronda

La expresión de "ronda" se refiere a los ciclos de la encuesta. Por ejemplo, si una encuesta continua es trimestral, entonces cada tres meses se obtendrán resultados de la encuesta. Por lo tanto una ronda de la encuesta es equivalente a un trimestre.

2.3 Puntos importantes para el diseño de encuestas

Cuando se realiza una encuesta es necesario determinar lo siguiente:

a. Objetivos

En todo diseño de encuesta, lo primero que tiene que ir definido son los objetivos que se desean lograr con la realización de la encuesta. Estos determinaran en gran manera todo lo siguiente.

b. Ámbito de la encuesta

El ámbito de una encuesta se puede desglosar en tres apartados que son los siguientes:

Ámbito poblacional

El ámbito poblacional es la especificación de la población a la que está dirigida la encuesta. También se puede incluir las exclusiones de la población según objetivos de la encuesta.

Ámbito geográfico

El ámbito geográfico está definido como el territorio geográfico en el cual se realizarán las inferencias a partir de la encuesta realizada.

Ámbito temporal

En el ámbito temporal se especifica el periodo de referencia de los resultados de la encuesta y también el periodo de referencia de la información recogida.

c. Marco muestral

El marco muestral es un listado que permite identificar a cada elemento de una población para asignar una probabilidad de selección a cada unidad muestral. Sin embargo, en muchas encuestas, especialmente en encuestas a hogares realizadas por los institutos de estadística, no se cuenta con un marco muestral a nivel individual, por lo que se recurre a otro tipo de marcos muestrales. Generalmente estos marcos muestrales son marcos de áreas, que son conglomerados de personas o viviendas.

d. Diseño muestral

El diseño muestral es un apartado de suma importancia para la realización de cualquier encuesta, ya que a partir de este se puede determinar la confiabilidad de los resultados obtenidos.

El diseño muestral debe contener lo siguiente:

- Tipo de muestreo
- Estratificación de las unidades de muestreo (si corresponde)

- Tamaño de muestra
- Afijación de la muestra (en caso de un muestreo estratificado)
- Tipo de selección de la muestra

En el caso de diseños de encuestas tipo panel rotatorio además de los anteriores puntos citados, debe también ir incluidos los siguientes puntos:

- Modelo de rotación de la muestra
- Distribución de la muestra en el tiempo

e. Estimadores

Los estimadores que se utilizarán en la encuesta deben ir descritos en esta parte. Existen estimadores de razón, estimadores de promedio, totales, etc. Cada estimador definido debe contar el error muestral respectivo calculado de acorde al diseño muestral adoptado en la encuesta.

2.4 Encuestas Longitudinales

De acuerdo Lynn (2005,p.3), una encuesta longitudinal es aquella que reúne datos de una muestra en múltiples ocasiones a través del tiempo. Sin embargo no confundamos una encuesta longitudinal con datos longitudinales, puesto que datos longitudinales se remiten a los mismos elementos muestrales pero no necesariamente recolectados en diferentes ocasiones, sino que pueden haber sido recogidos mediante la memoria retrospectiva o mediante un recopilado de información proveniente de registros. Por lo tanto, lo que diferencia las encuestas longitudinales de los datos longitudinales es el proceso de recogida de datos, aunque las encuestas longitudinales generan, por supuesto, datos longitudinales.

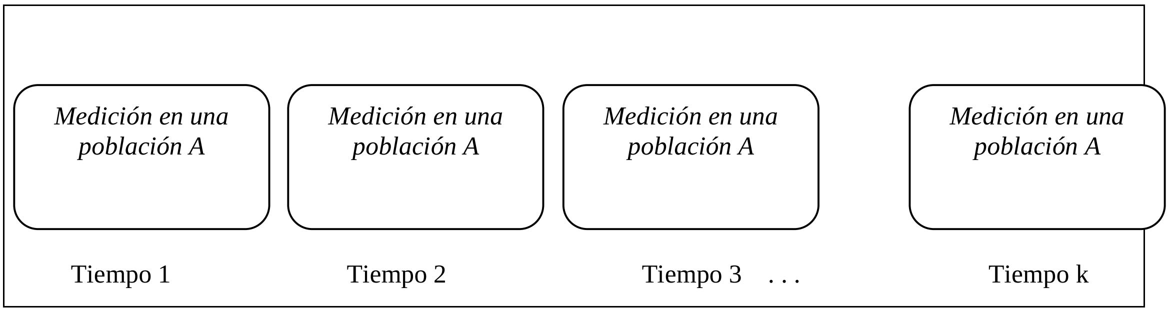
Las encuestas longitudinales se dividen en tres:

- Encuestas longitudinales de tendencia
- Encuestas longitudinales de evolución de grupo
- Encuestas longitudinales panel

2.4.1 Encuesta Longitudinal de Tendencias

Las encuestas con diseño de tendencias o "trend" son aquellas que analizan cambios de alguna característica o variable de interés en varios puntos del tiempo. Este estudio se realiza a una población en general. Por lo tanto la característica distintiva de estos diseños de tendencia es que la atención se enfoca en una población.

Figura 2.1: Diseño longitudinal de tendencias[1]



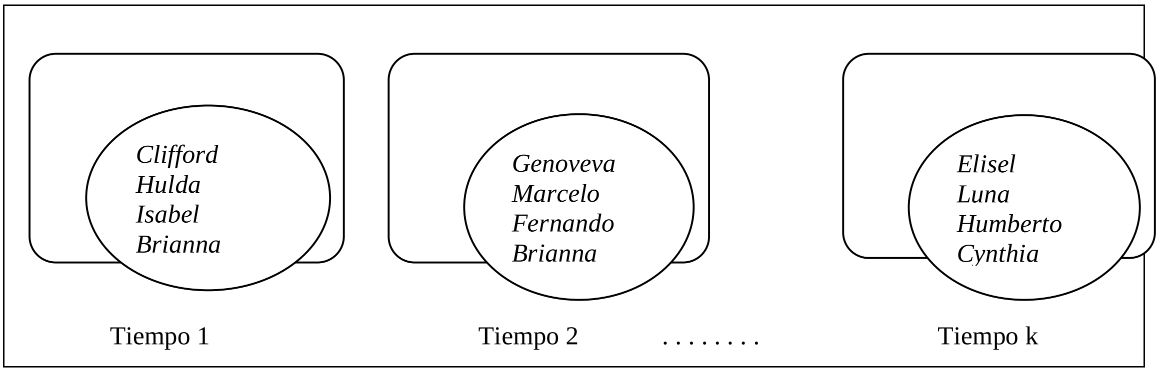
2.4.2 Encuesta longitudinal de evolución de grupo

Las encuestas con diseño de evolución de grupo o también llamados estudios "*cohort*", examinan cambios en varios puntos del tiempo en sub poblaciones o grupos específicos. Para definir un grupo de este tipo, se puede utilizar un criterio temporal o algún otro de interés, de modo que todos los elementos tengan una característica similar.

[1] Figura extraída de www.tecnicas-de-estudio.org

En este tipo de diseños se extrae una muestra cada vez que se mide la sub población, teniendo así muestras independientes dentro la sub población toda vez que se realice el estudio.

Figura 2.2: Diseño longitudinal de evolución de grupo



Esto quiere decir que algunos o todos los sujetos pueden cambiar, pero la sub población sigue siendo la misma.

La diferencia entre diseño de tendencia y de evolución de grupo se ve en el siguiente ejemplo: "Un investigador está interesado en estudiar las actitudes de los maestros respecto a las asociaciones de profesionales en la Región "A". Las actitudes son medidas cada tres años durante un periodo de 15 años. En cada momento que se hace la medición, se selecciona de la población de maestros existente en ese momento, una muestra de ellos. La membresía de la población puede cambiar a través del tiempo al menos parcialmente (algunos pueden dejar de ser maestros o ingresar nuevos maestros), pero en cualquier momento o tiempo la población es la misma: los maestros de la Región 'A" (llamada población general). Éste sería un ejemplo de un diseño de tendencia.

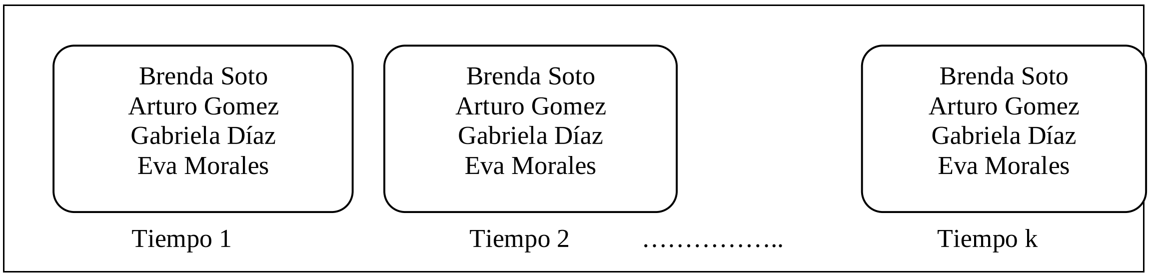
Si el investigador estuviera interesado en estudiar las actitudes hacia los sindicatos de profesionales por parte de los maestros que se iniciaron como tales en 1986, en la Región "A", el estudio involucraría el análisis de una sub población o grupo específico.

Tres años después, la siguiente muestra se obtendría de lo que queda de esa sub población, la cual en 1989 estará constituida por maestros con tres años de experiencia. Desde luego, algunos de los maestros que se iniciaron como tales en 1986 habrán dejado la docencia, y el estudio incluirá sólo las actitudes del grupo o sub población de maestros que comenzaron a serlo en dicho año y en 1989 continúan en el magisterio (de toda la población de maestros se estudia a una sub población). Éste sería un ejemplo de diseño de evolución de grupo o cohort.” (Wiersma, 1986, p. 208 citado en Lynn P. 2005)

2.4.3 ¿Qué es una encuesta de panel?

Una encuesta panel es aquella que sigue y hace mediciones sucesivas sobre un mismo grupo de observaciones en distintos momentos del tiempo. Este tipo de encuesta nos ayuda a conocer los cambios que experimentan los individuos, con relación a distintas variables de interés.

Figura 2.3: Diseño longitudinal panel[2]



2.4.3.1 Puntos fuertes en la recogida de datos

Lynn P. (2005, p.5) cita algunos puntos fuertes en la recogida de datos que son:

[2] Figura extraída de www.tecnicas-de-estudio.org

- Es posible reunir historias continuadas mucho *más largas* sobre sucesos y transiciones que las que se podrían reunir retrospectivamente en una sola entrevista. Esto quiere decir que se puede recoger datos longitudinales mediante una encuesta no longitudinal, apelando a los recuerdos del informante.

- Es posible reunir datos más *precisos* de los que sería posible en una sola entrevista con recuerdos retrospectivos, en los cuales los datos podrían estar sujetos a importantes errores de memoria.

- Es posible reunir información sobre expectativas y alternativas que no estén afectadas por sucesos y resultados posteriores; también sobre los sucesos y resultados posteriores de las mismas unidades muestrales.

2.4.3.2 Puntos fuertes en el análisis de datos

Cuando se elige trabajar con un determinado tipo de encuesta, esta decisión está relacionada con el tipo de análisis que se quiere lograr. Por lo tanto no se puede decir que las encuestas longitudinales son mejores que las encuestas de otro tipo, puesto que cada una tiene una razón para su uso. Sin embargo algunos análisis se pueden realizar solamente mediante encuestas longitudinales. Entonces se puede decir las que ventajas son estrictamente analíticas e incluyen los siguientes aspectos:

- El análisis del cambio bruto
- El análisis del cambio medio a nivel de unidad
- El análisis de la estabilidad o inestabilidad en cuanto a características;
- El análisis de las características en términos de tiempo de los sucesos o circunstancias, como son la frecuencia, el momento y la duración;
- Análisis de la naturaleza ordinal de los sucesos, lo cual sirve, a menudo, para clarificar problemas de causalidad.

2.4.3.3 Puntos débiles en la recogida de datos

Existen dos aspectos importantes en las encuestas panel que son potencialmente perjudiciales para la recolección de datos:

- Condicionamiento de paneles
- Desgaste de los paneles, en inglés también conocido como "attrition"

El condicionamiento de panel se refiere a la posibilidad de que una persona que haya participado con anterioridad en la encuesta proporcione respuestas distintas que como si participara por primera vez.

El desgaste de los paneles se refiere a la continua pérdida de muestra conforme transcurren los ciclos de entrevista. Este desgaste puede deberse a la negativa de participación en la encuesta, etc.

2.4.3.4 Puntos débiles en el análisis de datos

Muchas veces las encuestas longitudinales tipo panel no son tan adecuadas como las encuestas transversales cuando se desea obtener estimaciones transversales. Específicamente si las estimaciones transversales que se desea calcular son obtenidas mediante un promedio de dos o más rondas consecutivas, las encuestas longitudinales generan varianzas más grandes que las varianzas generadas con muestras independientes. Esto puede parecer un punto débil, pero simplemente se trata de que las encuestas panel no han sido diseñadas para ello (Longitudinal surveys Methodology, p. 11).

Veamos lo que pasa cuando se desea estimar un promedio:

Sea:

$\hat{t}_1$: Estimación en el tiempo 1

$\hat{t}_2$: Estimación en el tiempo 2

Donde el estimador del promedio se calcula:

$$\hat{\theta}_p = \frac{\hat{t}_1 + \hat{t}_2}{2} \tag{2.1}$$

La varianza tiene la siguiente forma:

$$Var\left[(\hat{t}_1 + \hat{t}_2)/2\right] = \frac{1}{4}\left[V(\hat{t}_1) + V(\hat{t}_2) + 2\operatorname{cov}(\hat{t}_1, \hat{t}_2)\right] \tag{2.2}$$

Es posible calcular la covarianza entre las dos rondas porque estamos trabajando con muestras panel.

Si en vez de dos muestras panel se tiene dos muestras independientes la varianza para el promedio se calcula de la siguiente manera:

$$Var\left[(\hat{t}_1 + \hat{t}_2)/2\right] = \frac{1}{4}\left[V(\hat{t}_1) + V(\hat{t}_2)\right] \tag{2.3}$$

La covarianza para dos muestras independientes es cero. Comparando la ecuación 2.2 y la ecuación 2.3 concluimos lo siguiente:

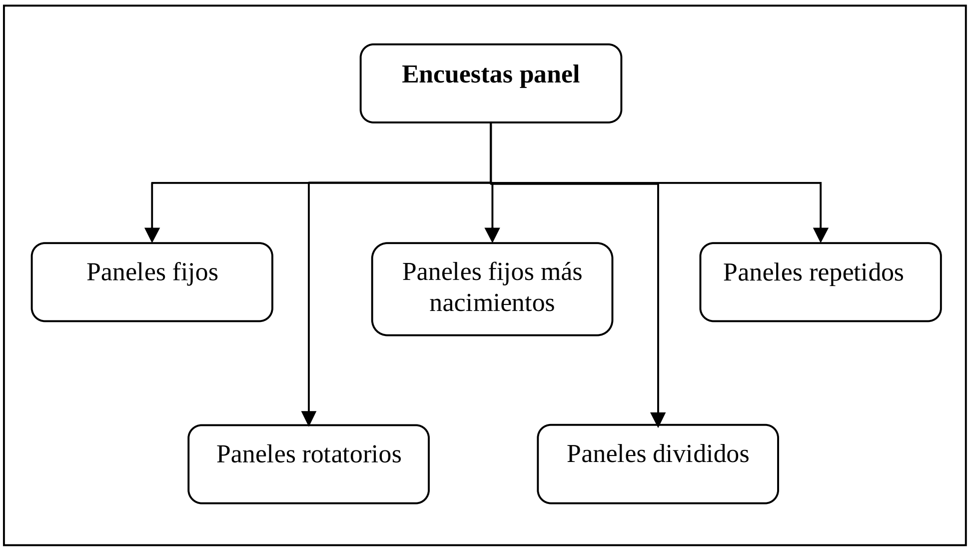
- Si la covarianza es positiva, el mejor estimador para el promedio se obtiene mediante dos muestras independientes, ya que la varianza del estimador del promedio obtenido con muestras independientes será menor que el que se obtiene con muestras dependientes.
- Si la covarianza es negativa, entonces ocurre que el mejor estimador del promedio se obtiene mediante muestras dependientes.

La ocurrencia de este segundo caso es muy poco probable, ya que estamos hablando de una misma variable en tiempos distintos. En el capítulo siguiente desarrollaremos con más detalle los estimadores de promedio con sus respectivas varianzas.

2.4.4 Clasificación de las Encuestas Panel

Las encuestas panel pueden clasificarse en 5 grandes categorías:

Figura 2.4: Clasificación de las encuestas panel

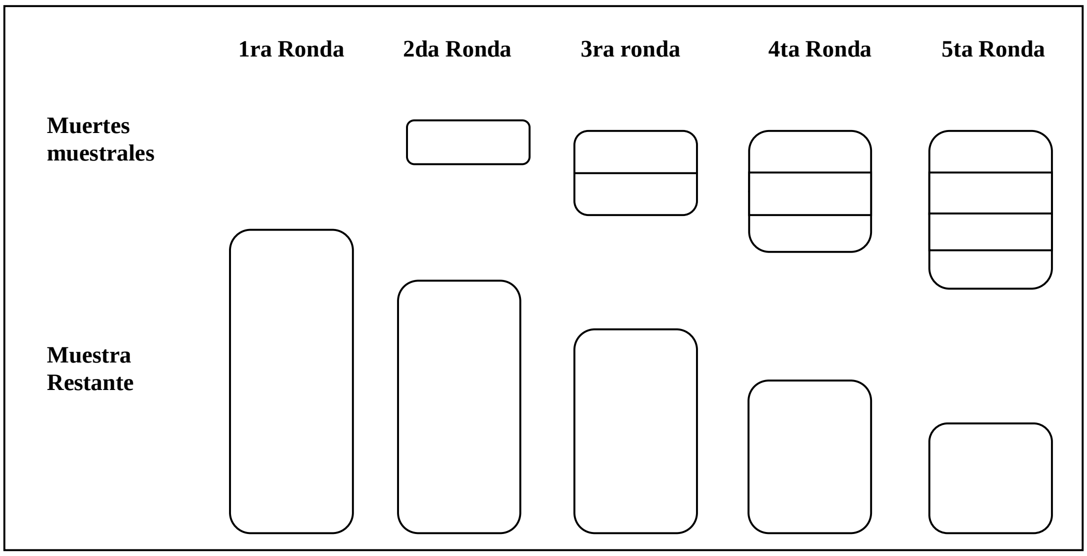


2.4.4.1 Paneles Fijos

Este tipo de panel recoge datos de las encuestas de las mismas unidades en múltiples ocasiones. Después de la selección inicial de la muestra ya no se pueden agregar modificaciones. En principio, la única perdida para la muestra es a través de las "muertes" de la población. Esto ocurre cuando se elige una muestra de la población de interés para tomar parte en la primera ronda. Poco después, tendrá lugar la segunda ronda de la encuesta, y para entonces dichas unidades ya no están en la población de interés, y por ello ya no serán parte de la encuesta. A esas unidades no referimos como muertes muestrales.

Por lo tanto, la muestra para la encuesta en la segunda ronda será menor o igual que en la primera ronda. El tiempo transcurrido entre rondas y la tasa de muerte muestral será factor importante para la reducción de la muestra.

Figura 2.5: Diseño panel fijo[3]



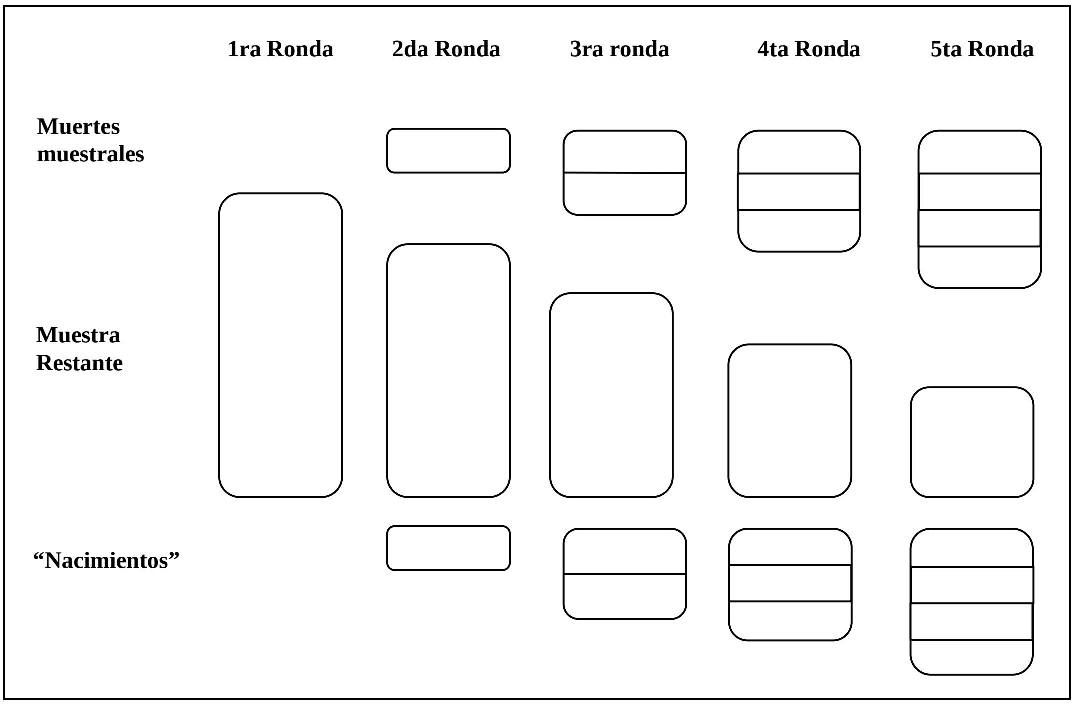
[3] Gráfico extraído de Longitudinal surveys methodology, p. 18

2.4.4.2 Paneles fijos más nacimientos

El panel fijo más nacimientos es muy similar al caso anterior, con la diferencia de que ahora se añade a la muestra de la población muestras regulares de los nacimientos recientes. Esto hace que se convierta en una ventaja, puesto que con este tipo de diseño la muestra de cada ciclo representa la población transversal actual, facilitando que realicen estimaciones transversales en paralelo con las estimaciones longitudinales.

Una gran mayoría de las encuestas panel a hogares tienen este diseño, dado que en cada ciclo se agrega una muestra de los nacimientos, y con esto se logran dar resultados transversales y longitudinales.

Figura 2.6: Diseño panel fijo más nacimientos[4]



[4] Gráfico extraído de “Longitudinal surveys methodology”, p. 19

Observando la figura 2.6 se puede concluir que el tamaño de la muestra para las rondas posteriores se mantendrá más o menos constante, puesto que las muertes muestrales serán sustituidas por los nacimientos. Esto facilita la posibilidad de realizar estimaciones transversales con un control previo del comportamiento de unidades muestrales en las muertes y nacimientos muestrales.

2.4.4.3 Paneles repetidos

Este diseño consta de una serie de encuestas panel, que pueden o no superponerse en el tiempo. Normalmente, se diseña cada panel para representar una población *equivalente*, es decir, la misma definición de población aplicada en un punto del tiempo diferente. Es frecuente encontrar este diseño en las encuestas para las personas que terminan el colegio, egresados como titulados de la universidad. Cada panel podría consistir en una muestra de una cohorte en particular de un año de edad elegida en años diferentes, constando cada panel de por lo menos tres ciclos en los últimos tres años. (Longitudinal surveys Methodology, p. 18)

2.4.4.4 Paneles Rotatorios

Cuando se trata del diseño de un panel rotativo, se sustituyen proporciones predeterminadas de unidades muestrales en cada ocasión del trabajo de campo. Normalmente, cada unidad permanecerá en la muestra durante el mismo número de rondas. Se ha utilizado una gran variedad de patrones de rotación. Las encuestas que utilizan diseños de panel rotatorio recogen, por lo general, los mismos datos de cada unidad en cada ronda, con el fin de permitir dicha combinación.

Los diseños de panel rotatorio se usan a menudo cuando los objetivos principales son estimaciones transversales y estimaciones a corto plazo del cambio neto y bruto. En muchos países, las encuestas sobre empleo utilizan un diseño de panel rotatorio. (Longitudinal surveys Methodology, p. 19).

2.4.4.5 Paneles Divididos (Split Panel)

Este diseño de panel incluye una combinación de muestras transversales y muestras panel en cada ocasión que se realiza la encuesta. Este tipo de diseño fué descrito por Kish (1987, p.181-183, citado en Lynn p., 2005) consiste en una muestra de panel fijo de donde se recogen datos de cada ocasión más una muestra transversal complementaria en cada ocasión.

Una serie de encuestas transversales en las cuales una proporción de los elementos de muestra es deliberadamente retenida en la muestra para encuestas consecutivas se denomina también encuesta de superposición a la que se puede considerar como un tipo de panel dividido. Normalmente se toman en cuenta estos diseños cuando lo que se busca es realizar estimaciones, tanto de transversales como longitudinales, o también en caso de las dudas antes mencionadas sobre la capacidad de que las muestras longitudinales puedan proporcionar estimaciones transversales adecuadas. (Longitudinal surveys Methodology, p. 11)

2.5 Encuestas Panel Rotatorio con Traslape Parcial

Dependiendo de los objetivos de la encuesta y los resultados que se deseen lograr, se puede combinar los diferentes diseños panel para alcanzar dichos objetivos. Esto quiere decir por ejemplo que podríamos combinar paneles rotatorios que consideren los nacimientos de la población para poder dar estimaciones transversales y longitudinales además de indicadores dinámicos.

Las muestras con rotación de paneles se pueden utilizar cuando una encuesta se repite cada año, trimestre o cada mes.

2.5.1 ¿Por qué una muestra con rotación de paneles?

Como ya mencionamos anteriormente, si el objetivo de una investigación es analizar los cambios de valor de algún parámetro de la población a través del tiempo, lo mejor diseñar muestras independientes. Sin embargo si además de indicadores transversales se desea calcular indicadores de cambios brutos y cambios entre rondas, entonces lo mejor es diseñar encuestas panel rotatorio con traslape parcial.

Las ventajas que se tienen con un diseño panel rotatorio son las siguientes:

- Estimar cambios (entre rondas de la encuesta) con mayor precisión que con muestras independientes. Esto gracias a que existe correlación entre rondas.
- Reducción del trabajo de campo, ya sea en listados, preguntas del cuestionario y tiempo de entrevista
- Mejorar las estimaciones de una ronda con información de una ronda anterior
- En cualquier encuesta panel la tasa no respuesta es siempre más alta. Sin embargo al tener una muestra panel rotatorio las unidades muestrales participan en la encuesta un número limitado de veces. Como consecuencia la tasa de no respuesta no es tan alta como si se trabajara con un diseño panel sin rotación, donde las unidades muestrales son estudiadas un número indefinido de veces

Algunas desventajas de este diseño son.

- Las estimaciones calculadas de tiempos agregados son menos precisas que las calculadas con muestras independientes. Cuando se trabaja con traslape parcial el efecto del traslape en la varianza del estimador será proporcional al porcentaje de traslape

- Surge problemas con salidas no planificadas de la muestra y por lo tanto se tiene que establecer procedimientos de reemplazo de esas unidades para no incrementar la tasa de no respuesta.

2.5.2 Nomenclatura de los Diseños Paneles

Al comenzar el primer capítulo se dieron algunas definiciones generales usadas en las encuestas panel. En esta parte se explicará otras definiciones adicionales con más detalle para una mejor comprensión.

2.5.2.1 Que es un Panel

Un panel es un grupo de unidades muestrales a las cuales se hace mediciones sucesivas por un tiempo o tiempo indefinido.

2.5.2.2 Definición del Periodo

El periodo o periodicidad de una encuesta está definida por la frecuencia de la obtención de resultados. Generalmente las encuestas a hogares que tienen un diseño panel tienen una periodicidad mensual o trimestral. A cada periodo de la encuesta nos referimos como rondas.

2.5.2.3 Paneles con visitas en periodos con descanso

Al momento del diseño de la encuesta se define el modelo de rotación de los paneles. Algunos modelos implican que cada panel tenga un número fijo de entrevistas, pero con periodos de descanso.
A estos modelos de rotación se los nombra: Panel, periodo de la encuesta y A – (B) – C.

donde:

- A+C = número de rondas
- B = periodos de descanso
- A+B+C = número de rondas de observación de una unidad muestal.

Veamos los siguientes casos de modelos de rotación con periodos de descanso.

Panel Trimestral 2 – (2) – 2

En este caso cada ronda dura un trimestre y los números 2 – (2) – 2 indican lo siguiente: Cada unidad muestral es entrevistada 2 trimestres consecutivos, después tienen un descanso de 2 trimestres y vuelven a ingresar a la encuestas por 2 últimos trimestres. Con este entendido, cada unidad muestral se entrevista un total de 4 veces.

Panel mensual 4 – (8) – 4

De modo similar al anterior caso, las rondas de este modelo tienen una duración de un mes. Cada unidad muestral es entrevistada un total de 8 veces; 4 meses consecutivos, descansa 8 meses para después ingresar nuevamente en los siguientes 4 meses.

2.5.2.4 Paneles con visitas en periodos consecutivos

A diferencia de los anteriores casos, también existen paneles con modelos de rotación que no contemplan periodos de descanso. A estos modelos se los nombra: Panel, periodo de la encuesta “X” donde X se define según el número de visitas que se realizará a cada unidad muestral. Veamos el siguiente modelo de rotación el cual no contempla periodos de descanso:

Panel trimestral 8

En este modelo de rotación cada unidad muestral es entrevistada 8 trimestres consecutivos, donde una ronda es igual a un trimestre. En este modelo no se contemplan rondas de descanso.

2.5.3 Tasa de Rotación

La tasa de rotación o también llamada la fracción de rotación es la determinación del número de rondas que participará una unidad muestral en la encuesta. Para determinar cuál debe ser la tasa de rotación se tiene que considerar lo siguiente:

- Si estimar los cambios es muy importante, entonces la tasa de rotación debe ser baja, es decir, cada unidad debe participar muchas rondas.

- La tasa de rotación tiene impacto directo en la tasa de no respuesta, puesto que el riesgo de pérdida de muestra aumenta cuando la tasa de rotación disminuye.

2.5.3.1 Modelos de Rotación

Existe una variedad de modelos de rotación utilizados en las diferentes encuestas[5] (ver Anexo 2.1 – 2.4) que realizan los institutos nacionales de estadística. En esta sección se cita algunos de ellos.

[5] Fichas metodológicas de encuestas panel en varios países

Modelo Panel 5

Este modelo de rotación divide la muestra en 5 paneles, esto permite que una unidad muestral participe 5 rondas consecutivas en la encuesta y después salga definitivamente.

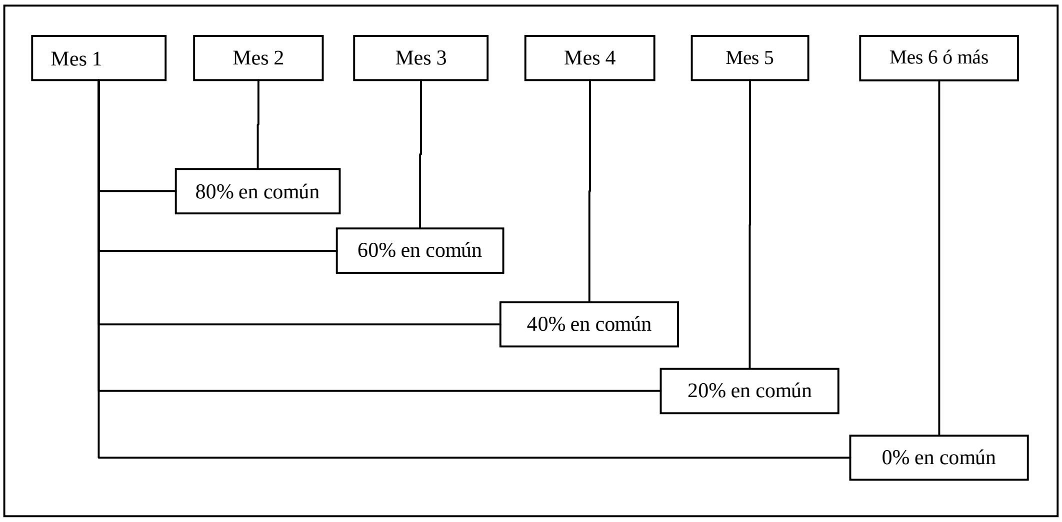
Uno de los países que utiliza este modelo de rotación es México, en la ENOE (Encuesta Nacional de Ocupación y Empleo). La encuesta tiene periodicidad mensual. El esquema de rotación se muestra en la figura 2.7.

Figura 2.7: Modelo de Rotación Panel 5

PANELES	1ra Ronda	2da Ronda	3ra Ronda	4ta Ronda	5ta Ronda	6ta Ronda	7ma Ronda
Panel 1	1ra visita						
Panel 2	1ra visita	2da visita					
Panel 3	1ra visita	2da visita	3ra visita				
Panel 4	1ra visita	2da visita	3ra visita	4ta visita			
Panel 5	1ra visita	2da visita	3ra visita	4ta visita	5ta visita		
Panel 6		1ra visita	2da visita	3ra visita	4ta visita	5ta visita	
Panel 7			1ra visita	2da visita	3ra visita	4ta visita	5ta visita
Panel 8				1ra visita	2da visita	3ra visita	4ta visita
Panel 9					1ra visita	2da visita	3ra visita
Panel 10						1ra visita	2da visita
Panel 11							1ra visita

El esquema de traslape se muestra en la figura 2.8

Figura 2.8 Traslape del esquema panel mensual 5



Modelo Panel 2 – (2) – 2

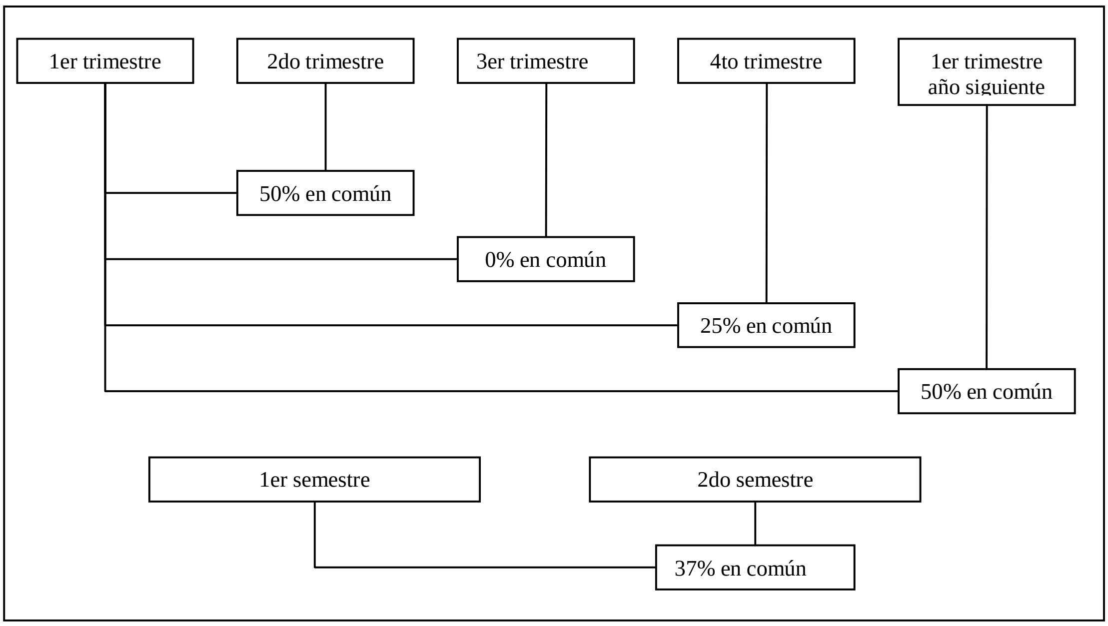
El modelo de rotación panel 2 – (2) – 2 es utilizado en algunos países de Latinoamérica. Por ejemplo en la Encuesta Permanente a Hogares de Argentina, también en la Encuesta de Empleo de Bolivia y anteriormente en la Encuesta de Ocupación y Desocupación del Gran Santiago, aunque actualmente cambió el modelo de rotación.

Según el esquema de rotación (ver figura 2.9), una unidad muestral es observada durante cuatro rondas, entra en la muestra dos rondas consecutivas, después descansa dos rondas consecutivas para después ingresar nuevamente en la muestra las siguientes dos rondas. Se ve claramente que para el inicio del modelo existen irregularidades en cuanto al número de visitas y los paneles. Por ejemplo los paneles 3 y 4 no se visitarán más de 2 rondas. Esto sucede hasta que se estabilice el modelo de rotación panel hasta una sexta ronda. A partir de la séptima ronda se introduce un 25% de muestra nueva y el resto es muestra antigua con una tasa de rotación constante.

Figura 2.9: Modelo de Rotación Panel 2 – (2) – 2

PANELES	1ra Ronda	2da Ronda	3ra Ronda	4ta Ronda	5ta Ronda	6ta Ronda	7ma Ronda	8va Ronda	9na Ronda	10ma Ronda	11ma Ronda	12ma Ronda
Panel 1	1ra visita	2da visita			3ra visita	4ta visita						
Panel 2	1ra visita			2da visita	3ra visita							
Panel 3	1ra visita	2da visita										
Panel 4	1ra visita											
Panel 5		1ra visita	2da visita			3ra visita	4ta visita					
Panel 6		1ra visita	2da visita									
Panel 7			1ra visita	2da visita			3ra visita	4ta visita				
Panel 8			1ra visita	2da visita								
Panel 9				1ra visita	2da visita			3ra visita	4ta visita			
Panel 10					1ra visita	2da visita			3ra visita	4ta visita		
Panel 11						1ra visita	2da visita			3ra visita	4ta visita	
Panel 12							1ra visita	2da visita			3ra visita	4ta visita

Figura 2.10: Traslape del esquema 2 – (2) – 2



Modelo Panel 4 – (8) – 4

Otro modelo de rotación es el 4 – (8) – 4. Este modelo es utilizado en encuestas con periodicidad mensual, como por ejemplo la Encuesta de Población Actual (CPS) desarrollado por la Oficina del Censo de Estados Unidos de América, bajo el esquema de rotación 4 – (8) – 4, que significa que un grupo de rotación permanece en la muestra durante 4 meses consecutivos, sale 8 meses y regresa para participar durante cuatro veces más. Es decir, en la Encuesta de Población Actual hecha mensualmente en los Estados Unidos una cuarta parte de las muestrales se reemplaza cada mes, de modo que un domicilio particular permanece en la muestra durante cuatro meses consecutivos. Ese domicilio se omite en la muestra por ocho meses, pero luego se re incluye por cuatro meses, incrementando así ligeramente la precisión de las comparaciones entre años consecutivos. Además de EEUU, otro país que también utiliza este modelo de rotación es Brazil, en la Encuesta Mensual de Empleo.

Las bondades de este modelo son:

- Varianzas estimadas mejoradas de diferencia entre meses, debido a que tres cuartos de la muestra se repiten entre meses consecutivos.
- Varianzas estimadas mejoradas de año a año porque la mitad de la muestra se repite en el mismo mes en dos años consecutivos.

Se puede observar en la Figura 2.11, que en un principio el esquema de rotación no es muy claro. Sin embargo a partir del mes 17 se estabiliza e ingresa solamente el 12.5% de muestra nueva y el resto es antigua. Los números dentro el cuadro indica que número de visita le corresponde a cada panel de rotación.

Figura 2.11: Modelo de rotación Panel 4 - (8) - 4

	p4	p3	p2	p1	p8	p7	p6	p5	p9	p10	p11	p12	p14	p15	p16	p17	p18	p19	p20	p21	p22	p23	p24	p25	p26	p27	p28	p29	p30	p31	p32	p33
mes 1	1	1	1	1	1	1	1	1																								
mes 2		2	2	2		2	2	2	1								1															
mes 3			3	3			3	3	2	1							2	1														
mes 4				4				4	3	2	1						3	2	1													
mes 5									4	3	2	1					4	3	2	1												
mes 6										4	3	2	1					4	3	2	1											
mes 7											4	3	2	1					4	3	2	1										
mes 8												4	3	2	1					4	3	2	1									
mes 9													4	3	2	1					4	3	2	1								
mes 10	2													4	3	2						4	3	2	1							
mes 11	3	3													2	3							4	3	2	1						
mes 12	4	4	4													4								4	3	2	1					
mes 13	5	5	5	5																					4	3	2	1				
mes 14		6	6	6													5									4	3	2	1			
mes 15			7	7													6	5									4	3	2	1		
mes 16				8													7	6	5									4	3	2	1	
mes 17																	8	7	6	5									4	3	2	1
mes 18																		8	7	6	5									4	3	2
mes 19																			8	7	6	5									4	3
mes 20																				8	7	6	5									4
mes 21																					8	7	6	5								

De la figura 2.11 se puede ver el traslape entre rondas es como muestra la Tabla 2.1.

Tabla 2.1

Intervalo de meses	Proporción común de la muestra (%)
1	75
2	50
3	25
4 – 8	0
9	12.5
10	25
11	37.5
12	50
13	37.5
14	25
15	12.5
16 y más	0

Modelo Panel 3 – (1) – 2

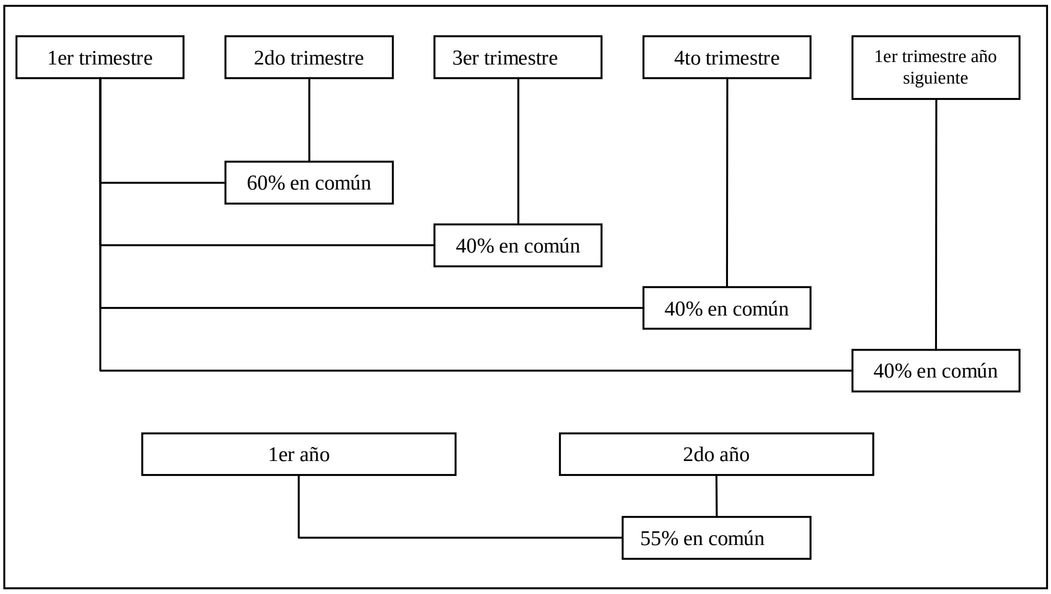
Se puede mencionar también otro modelo de rotación el cual consistente en el diseño de panel que proporciona datos para un análisis longitudinal. Este modelo tiene la característica de contar con un traslape entre dos años consecutivos igual a 60%. Con una encuesta trimestral la intención es guardar la alta coincidencia entre los años consecutivos y establecer también traslape alto para trimestres consecutivos.

Las unidades muestrales serán entrevistados tres trimestres consecutivos, serán omitidas por un trimestre e incluido otra vez por un período de 2 trimestres. Este modelo fue aplicado en Finlandia (Anexo 2.3).

Figura 2.12: Modelo de rotación panel 3 – (1) – (2)

PANELES	1ra Ronda	2da Ronda	3ra Ronda	4ta Ronda	5ta Ronda	6ta Ronda	7ma Ronda	8va Ronda	9na Ronda	10ma Ronda	11ma Ronda	12ma Ronda
Panel 1	1ra visita	2da visita	3ra visita		4ta visita	5ta visita						
Panel 2	1ra visita	2da visita		3ra visita	4ta visita							
Panel 3	1ra visita		2da visita	3ra visita								
Panel 4	1ra visita	2da visita										
Panel 5	1ra visita											
Panel 6		1ra visita	2da visita	3ra visita		4ta visita	5ta visita					
Panel 7		1ra visita	2da visita									
Panel 8			1ra visita	2da visita	3ra visita		4ta visita	5ta visita				
Panel 9				1ra visita	2da visita	3ra visita		4ta visita	5ta visita			
Panel 10					1ra visita	2da visita	3ra visita		4ta visita	5ta visita		
Panel 11						1ra visita	2da visita	3ra visita		4ta visita	5ta visita	
Panel 12							1ra visita	2da visita	3ra visita		4ta visita	5ta visita
Panel 13								1ra visita	2da visita	3ra visita		4ta visita

Figura 2.13: Traslape del esquema 3 – (1) – 2



2.5.4 Importancia de la estratificación

Al diseñar cualquier encuesta, una técnica muy extendida es la estratificación de de la población que se desea encuestar previa a la selección de la muestra. Esta sirve para clasificar a la población en subpoblaciones (estratos) basándose en información auxiliar conocida sobre toda la población.

Refiriéndose específicamente a encuestas panel rotatorio, la estratificación de la población es una opción para el diseño. Se debe controlar muy bien los paneles que salen de la muestra, de tal forma que el panel que entra en la muestra debe tener características similares al panel que terminó con todas las rondas planificadas. Por lo tanto, emplear la estratificación es reducir la probabilidad de que por mala fortuna se seleccione un número de unidades de muestreo desproporcionadamente grande (o pequeño) de una subpoblación que se considera importante para el análisis. La estratificación se aplica para asegurar la representación adecuada de los grupos de subpoblación importantes sin sesgar la operación de selección. De este modo se sugiere una estratificación del marco muestral, siendo esta no obligatoria, simplemente una opción a considerar en el diseño muestral.

2.5.4.1 Estratificación explicita

La estratificación explícita suele aplicarse en cada etapa de selección (si es que el diseño muestral es multietápico). No obstante, sus beneficios son especialmente notables en el muestreo de unidades muestrales de primera etapa. Por ello es muy importante estratificar previo a la selección.

La estratificación divide las unidades de la población en subgrupos mutuamente excluyentes y colectivamente exhaustivos. Luego, de cada estrato se selecciona muestras independientes. Uno de los objetivos primarios de la estratificación es mejorar la precisión de las estimaciones de las encuestas. En ese caso la conformación de los estratos debe ser tal que las unidades del mismo estrato sean lo más homogéneas posible, y las unidades de los diferentes estratos lo más heterogéneas posible con respecto a las características de interés para la encuesta.

A veces puede ser necesario usar varias variables en la estratificación: En tales casos debemos guiarnos por los siguientes factores: es preferible que las variables de estratificación no guarden relación entre sí, sino con la variable de la encuesta; no es preciso ser exhaustivo a la hora de crear las casillas (las más pequeñas o menos importantes pueden combinarse); en general, se otiene mayor beneficio empleando divisiones rudimentarias de varas variables que divisiones más afinadas de una sola variable. (Diseños de Muestras para Encuestas a Hogares, 2009)

2.5.4.2 Estratificación Implícita

Dentro de cada estrato explícito, al seleccionar las unidades muestrales se utiliza muchas veces una técnica conocida con el nombre de estratificación implícita. Esta

práctica será de gran utilidad en las encuestas panel rotatorio, ya que se tendrá un mejor control de las unidades muestrales que salen y entran en la muestra.
Para aplicar correctamente esta técnica, la estratificación implícita exige el uso de la selección sistemática en la primera etapa de muestreo. Realizar el procedimiento es muy sencillo e implica, en primer lugar, organizar el archivo de unidades primarias de muestreo en orden geográfico. En muchos países en las encuestas que realizan los institutos de estadística las zonas urbanas se ordenarán por provincias, y dentro de la provincia, por distrito, y lo mismo en las zonas rurales. El siguiente paso consiste en seleccionar en forma sistemática las unidades muestrales del archivo ordenado. La selección sistemática se lleva a cabo bien a través de muestreo con igual probabilidad o muestreo con probabilidad proporcional al tamaño.

Entonces, una ventaja importante de la estratificación implícita es que elimina la necesidad de establecer estratos geográficos explícitos, lo cual suprime a la vez la necesidad de asignar muestra a dichos estratos, sobre todo si se emplea el muestreo proporcional. Otra ventaja es la simplicidad, dado que el método solo exige la ordenación del archivo y la aplicación del intervalo de muestreo.

2.5.5 Selección de la muestra y asignación a los paneles

Una vez definido el modelo de rotación se prosigue con la selección de la muestra. Para la primera ronda se selecciona la muestra como cualquier muestra independiente. Esto significa que es posible utilizar varios diseños. En el anterior párrafo vimos la importancia y las ventajas de que la muestra sea estratificada, y si el tamaño es importante se puede puede combinar con una selección proporcional al tamaño.
Cualquiera que fuese el diseño, la muestra debe ser dividida en tantos paneles que se hayan definido.

Algunas consideraciones importantes al momento de realizar la asignación de la muestra a los paneles son:

- Definir el número de paneles
- La asignación de las unidades muestrales a los paneles debe ser aleatoria
- Por tratarse de una encuesta longitudinal panel rotatorio la muestra debe asignarse en un específico tiempo (mes, semana) para cada ronda de la encuesta
- El número de unidades muestrales de un panel que ingresa en la muestra, está definida por el panel que sale de la encuesta. Es decir, el número de unidades muestrales en cada estrato será igual a las unidades muestrales que había en el panel que salió de la muestra.

Para comprender mejor el proceso de asignación a los paneles veamos una simulación de una muestra con el modelo de rotación 2-(2)-2.

Para el caso del panel trimestral 2 - (2) - 2 los paneles rotan de la siguiente manera:

Figura 2.14: Modelo de rotación panel trimestral 2 – (2) – 2

TRIMESTRE	PANELES			
1er Trimestre	P1	P2	P3	P4
2do Trimestre	P5	P1	P6	P3
3er Trimestre	P7	P5	P8	P6
4to Trimestre	P9	P7	P2	P8
1er Trimestre	P10	P9	P1	P2
2do Trimestre	P11	P10	P5	P1
3er Trimestre	P12	P11	P7	P5
4to Trimestre	P13	P12	P9	P7
1er Trimestre	P14	P13	P10	P9
2do Trimestre	P15	P14	P11	P10
3er Trimestre	P16	P15	P12	P11
4to Trimestre	P17	P16	P13	P12

Los colores pueden ayudar significativamente a reemplazar los paneles nuevos que ingresarán en la muestra en lugar de los paneles que ya cumplieron el número de rondas previstas y salieron de la muestra. En el caso del 4to trimestre el panel que ingresa en la muestra es el P9 y el panel que sale de la muestra es P6, sin embargo P9 tiene que tener las características de P5. Eso lo vemos gracias a los colores que indican el panel faltante.

Tomaremos como ejemplo el supuesto caso de una muestra de tamaño 200 seleccionada de 3 estratos independientes.

Sea la distribución

Estrato 1: 64 unidades muestrales

Estrato 2: 69 unidades muestrales

Estrato 3: 67 unidades muestrales

Cada unidad de muestreo seleccionada será asignada a uno de los paneles de la primera ronda P1, P2, P3, y P4. Esto se realiza con una lista de las unidades muestrales seleccionadas en orden de los estratos. Comenzando con un panel aleatorio entre los cuatro paneles (por ejemplo P2, P4, P1 y P3), la primera unidad muestral es asignada a P2, la segunda a P4, la tercera a P1, etc. En un nuevo estrato se continúa con la secuencia. Para la asignación de las semanas se procede de la misma manera que para los paneles.

Después de realizar la asignación la muestra queda como muestra la Tabla 2.2.

Tabla 2.2
Distribución de la muestra por Panel y Semana de Referencia

SEMANA	PANEL				Total
	P1	P2	P3	P4	
1	4	4	3	4	**15**
2	4	3	4	4	**15**
3	3	4	4	4	**15**
4	4	4	4	3	**15**
5	4	4	4	4	**16**
6	4	4	4	4	**16**
7	4	4	4	4	**16**
8	4	4	4	4	**16**
9	4	4	4	4	**16**
10	4	4	3	4	**15**
11	4	3	4	4	**15**
12	3	4	4	4	**15**
13	4	4	4	3	**15**
Total	**50**	**50**	**50**	**50**	**200**

Tabla 2.3
Distribución de la muestra por Panel y Semana de Referencia

Semana trimestral	Estrato 1				Total	Estrato 2				Total	Estrato 3				Total
	P1	P2	P3	P4		P1	P2	P3	P4		P1	P2	P3	P4	
1	1	2	1	1	5	2	1	1	1	5	1	1	1	2	5
2	2	1	1	1	5	1	1	1	2	5	1	1	2	1	5
3	1	1	1	2	5	1	1	2	1	5	1	2	1	1	5
4	1	1	1	1	4	1	2	2	1	6	2	1	1	1	5
5	1	1	2	1	5	2	2	1	1	6	1	1	1	2	5
6	1	2	1	1	5	2	1	1	2	6	1	1	2	1	5
7	2	1	1	1	5	1	1	2	2	6	1	2	1	1	5
8	1	1	1	2	5	1	1	2	1	5	2	2	1	1	6
9	1	1	2	1	5	1	2	1	1	5	2	1	1	2	6
10	1	2	1	1	5	2	1	1	1	5	1	1	1	2	5
11	2	1	1	1	5	1	1	1	2	5	1	1	2	1	5
12	1	1	1	2	5	1	1	2	1	5	1	2	1	1	5
13	1	1	2	1	5	1	2	1	1	5	2	1	1	1	5
Total	**16**	**16**	**16**	**16**	**64**	**17**	**17**	**18**	**17**	**69**	**17**	**17**	**16**	**17**	**67**

Tabla 2.4
Distribución de la muestra por Panel y Estrato

ESTRATOS	Panel				Total
	P1	P2	P3	P4	
Estrato 1	16	16	16	16	**64**
Estrato 2	17	17	18	17	**69**
Estrato 3	17	17	16	17	**67**
Total	**50**	**50**	**50**	**50**	**200**

Se puede observar que el balanceado se ajusta a los estratos, semanas de referencia y paneles. Nótese que se tiene 4 esquemas (Tabla 2.3) para cada estrato, los cuales deben mantenerse durante todas las rondas de la encuesta. Esto quiere decir que si por ejemplo nos encontramos en la segunda ronda donde ingresa el P5 este tendrá que tener las mismas características (número de unidades muestales por estrato, por semana de referencia) de P4 que fué el que salió de la muestra una ronda anterior.

Tabla 2.5
Distribución de la muestra P5 por Semana de Referencia según estrato

Semana	P5			
	Estrato 1	Estrato 2	Estrato 3	Total
1	1	1	2	4
2	1	2	1	4
3	2	1	1	4
4	1	1	1	3
5	1	1	2	4
6	1	2	1	4
7	1	2	1	4
8	2	1	1	4
9	1	1	2	4
10	1	1	2	4
11	1	2	1	4
12	2	1	1	4
13	1	1	1	3
Total	**16**	**17**	**17**	**50**

De esta forma se mantiene el mismo esquema para todas las rondas posteriores siempre y cuando se reemplacen los paneles apropiados. De lo contrario, a medida que pasan las rondas, el número de unidades muestrales por semana de referencia, estrato y paneles, pueden variar y como consecuencia inmediata se podrían tener serios problemas con el balanceado de la muestra.

Capítulo 3

Estimador de cambio, promedio y estimador combinado

3.1 Introducción

En el capítulo 2, se trató la metodología para el diseño panel de una encuesta. En el presente capítulo se verán tres tipos de estimadores que se pueden calcular mediante las encuestas tipo panel rotatorio con traslape parcial, estos son: Estimador de cambio de dos rondas distintas, estimador del promedio de dos o más rondas consecutivas y finalmente el estimador combinado. Dentro del desarrollo de cada uno de los estimadores se desarrolla también la varianza de los mismos.

3.2 Estimadores

3.2.1 Estimador de cambio de dos rondas consecutivas

Para la estimación puntual de una ronda la muestra funciona como una muestra independiente y las correlaciones entre rondas no se consideran. Sin embargo, para estimar cambios entre rondas se debe tomar en cuenta las correlaciones.

Sea el estimador de cambio entre dos ocasiones consecutivas:

$$\hat{\theta} = \hat{t}_{i+1} - \hat{t}_i \tag{3.1}$$

donde:

$\hat{t}_1$: Estimación en el tiempo 1

$\hat{t}_2$: Estimación en el tiempo 2

La varianza del estimador es:

$$\begin{aligned} Var(\hat{\theta}) &= Var(\hat{t}_{i+1} - \hat{t}_i) \\ Var(\hat{\theta}) &= Var(\hat{t}_{i+1}) + Var(\hat{t}_i) - 2Cov(\hat{t}_{i+1}, \hat{t}_i) \end{aligned} \tag{3.2}$$

Si $\hat{t}_i = \sum w_k x_k^i$, y la proporción de traslape de dos rondas consecutivas es $\beta_{i.i+1}$, el cálculo de la covarianza es:

$$\begin{aligned} Cov(\hat{t}_{i+1}, \hat{t}_i) = Cov(\textstyle\sum_1^{\beta_{i,i+1}n} w_k x_k^{i+1} + \sum_{\beta_{i,i+1}n+1}^{n} \quad , \sum_1^{\beta_{i,i+1}n} w_k x_k^i + \sum_1^{\beta_{i,i+1}n} \quad) \\ = \textstyle\sum_1^{\beta_{i,i+1}n} Cov(w_k x_k^{i+1}, w_k x_k^i) + 0 \\ = \textstyle\sum_1^{\beta_{i,i+1}n} w_k^2 Cov(x_k^{i+1}, x_k^i) \\ = \textstyle\sum_1^{\beta_{i,i+1}n} w_k^2 \rho_{i,i+1} \sqrt{Var(x_k^{i+1}) Var(x_k^i)} \end{aligned}$$

Si se asume varianza igual entre rondas:

$$Var(x_k^i) = Var(x_k^{i+1}) = Var(x_k) \tag{3.3}$$

$$Cov(\hat{t}_{i+1}, \hat{t}_i) = \frac{\sum_1^{\beta_{i,i+1}n} w_k^2}{\sum_1^n w_k^2} \rho_{i,i+1} Var(\sum_1^n w_k x_k)$$

$$Cov(\hat{t}_{i+1}, \hat{t}_i) = \beta_{i,i+1} \cdot \rho_{i,i+1} \cdot Var(\hat{t}) \tag{3.4}$$

donde:

$\rho_{i,i+1}$: Coeficiente de correlación entre la ronda *i* e *i+1*

Reemplazando (3.4) en la ecuación (3.2)

$$\begin{aligned} Var(\hat{\theta}) &= Var(\hat{t}_{i+1}) + Var(\hat{t}_i) - 2\beta_{i,i+1}\rho_{i,i+1}\sqrt{Var(\hat{t}_i)Var(\hat{t}_{i+1})} \\ &= 2Var(\hat{t}) - 2\beta_{i,i+1}\rho_{i,i+1}\sqrt{Var(\hat{t})Var(\hat{t})} \\ &= 2Var(\hat{t}) - 2\beta_{i,i+1}\rho_{i,i+1}\sqrt{Var^2(\hat{t})} \\ &= 2Var(\hat{t}) - 2\beta_{i,i+1}\rho_{i,i+1}Var(\hat{t}) \end{aligned}$$

$$Var(\hat{\theta}) = 2Var(\hat{t})(1 - \beta_{i,i+1}\rho_{i,i+1}) \tag{3.5}$$

Si asumimos una tasa de rotación es de 100% y $\beta_{i,i+1} = 0$ (muestras independientes), la ecuación (3.2) se reduce a:

$$\begin{aligned} Var(\hat{\theta}) &= Var(\hat{t}_{i+1}) + Var(\hat{t}_i) \\ Var(\hat{\theta}) &= 2Var(\hat{t}) \end{aligned} \tag{3.6}$$

Ahora se puede comprobar la eficiencia de la varianza obtenida mediante muestras panel rotatorio con una muestra independiente de la siguiente forma:

$$\frac{Var_{con.rot}(\hat{\theta})}{Var_{\sin.rot}(\hat{\theta})} = \frac{2Var(\hat{t})(1-\beta_{i,i+1}\rho_{i,i+1})}{2Var(\hat{t})} \tag{3.7}$$

Se ve que la varianza para la diferencia de dos rondas consecutivas es igual al error de una ronda multiplicado por $(1-\beta_{i,i+1}\rho_{i,i+1})$. Entonces veamos la reducción de la varianza con relación al caso de muestras independientes.

$$0 < \frac{Var_{con.rot}(\hat{\theta})}{Var_{\sin.rot}(\hat{\theta})} < 1 \qquad si: \qquad 1-\beta_{i,i+1}\rho_{i,i+1} < 1 \tag{3.8}$$

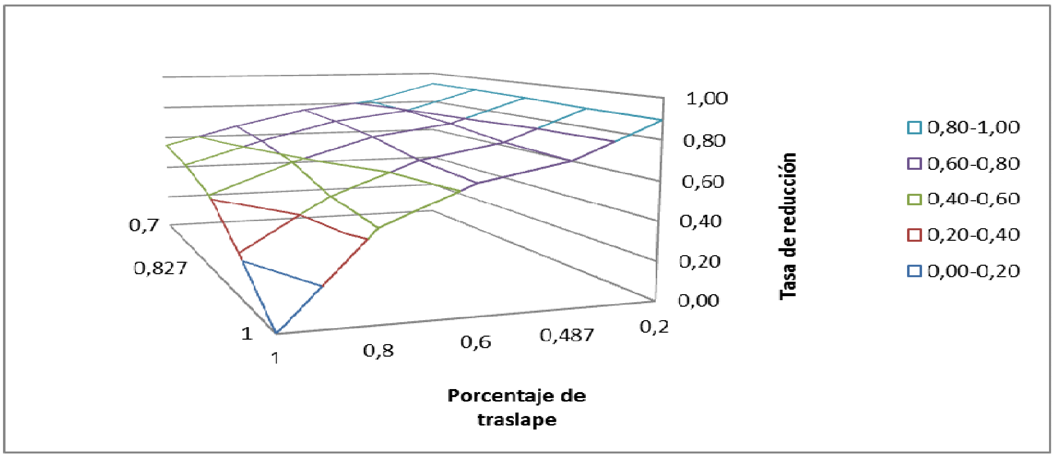
De la ecuación (3.8) se puede concluir que una muestra con rotación de paneles tendrá menor varianza que una muestra independiente si el coeficiente de correlación cumple con $\rho_{1,2}$ >0.

Veamos algunos valores de la tasa de reducción según traslape y coeficiente de correlación.

Tabla 3.1
Razón de la varianza con muestras panel y varianzas con muestras independientes
Estimador de cambio

Traslape	Coeficiente de Correlación			
	0.95	**0.85**	**0.75**	**0.7**
0.2	0.900	0.911	0.922	0.927
0.4	0.787	0.812	0.837	0.849
0.6	0.656	0.700	0.742	0.762
0.8	0.490	0.566	0.632	0.663
1	0.224	0.387	0.500	0.548

Figura 3.1: Razón de la varianza con muestras panel y varianzas con muestras independientes del Estimador de cambio



De la Tabla 3.1 podemos ver para varios casos de modelos de rotación. En el caso de encuestas panel 5 la varianza del estimador de la diferencia de dos rondas consecutivas tiene una reducción de 34% a 50% según el coeficiente de correlación. En cambio para el modelo 2 – (2) – 2 donde el porcentaje de traslape es de 50% la reducción de la varianza para el mismo estimador es de 15% a 25% aproximadamente.

Si el estimador del cambio es una prioridad entonces debe optarse por traslapes altos entre rondas y tasas de rotación pequeñas.

3.2.2 Estimador para el promedio

El promedio de dos o más rondas puede ser una opción para determinar un indicador de un periodo más amplio que de una sola ronda. En el caso de una encuesta panel trimestral se puede calcular el indicador anual mediante el promedio de cuatro trimestres consecutivos. También en el caso de una encuesta panel 5 se puede calcular el indicador anual mediante el promedio de 5 rondas consecutivas.

Al contrario del estimador de la diferencia, cuando se consideran promedios de 2 ó más rondas, se pierde precisión con muestras de rotación.

Veamos el promedio con dos rondas cualesquiera:

Sea el estimador del promedio entre dos ocasiones consecutivas:

$$\hat{\theta}_p = \frac{\hat{t}_i + \hat{t}_{i+1}}{2} \tag{3.9}$$

La varianza del estimador es:

$$Var(\hat{\theta}_p) = \frac{1}{4} Var(\hat{t}_{i+1} + \hat{t}_i)$$

$$Var(\hat{\theta}_p) = \frac{1}{4}(Var(\hat{t}_{i+1}) + Var(\hat{t}_i) + 2Cov(\hat{t}_{i+1}, \hat{t}_i)) \tag{3.10}$$

Asumiendo varianzas iguales en cualquier ronda y reemplazando la ecuación (3.4) en (3.10) se tiene:

$$\begin{aligned} Var(\hat{\theta}_p) &= \frac{1}{4}\left[Var(\hat{t}_{i+1}) + Var(\hat{t}_i) + 2\beta_{i,i+1}\rho_{i,i+1}\sqrt{Var(\hat{t}_i)Var(\hat{t}_{i+1})}\right] \\ &= \frac{1}{4}\left[2Var(\hat{t}) + 2\beta_{i,i+1}\rho_{i,i+1}\sqrt{Var(\hat{t})Var(\hat{t})}\right] \\ &= \frac{1}{2}\left[Var(\hat{t}) + \beta_{i,i+1}\rho_{i,i+1}Var(\hat{t})\right] \end{aligned}$$

$$Var(\hat{\theta}) = \frac{1}{2} Var(\hat{t})(1 + \beta_{i,i+1}\rho_{i,i+1}) \tag{3.11}$$

Analizando la razón de la varianzas de una muestra con rotación de paneles y una muestra independiente se tiene:

$$\frac{Var_{con.rot}(\hat{\theta})}{Var_{\sin.rot}(\hat{\theta})}=\frac{\frac{1}{2}Var(\hat{t})(1+\beta_{i,i+1}\rho_{i,i+1})}{\frac{1}{2}Var(\hat{t})}=1+\beta_{i,i+1}\rho_{i,i+1} \tag{3.12}$$

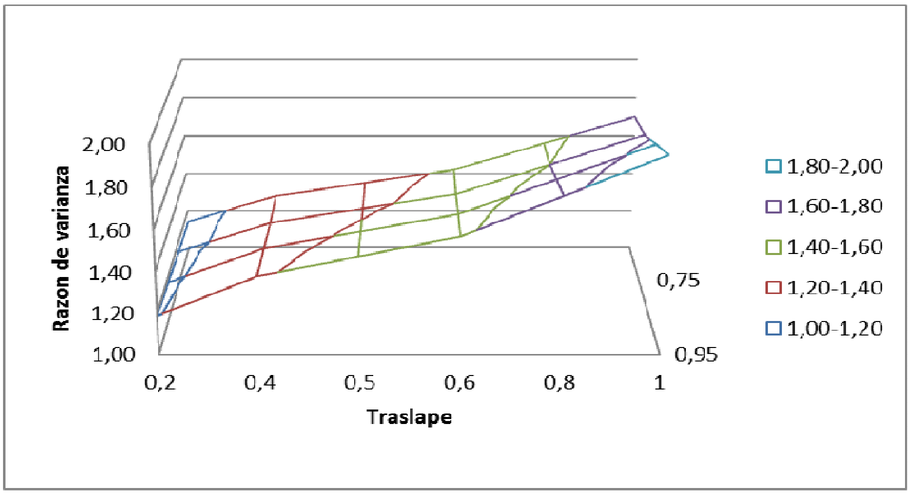
Es un hecho que la varianza con rotación de paneles es mayor o igual que la varianza obtenida mediante muestras independientes.

Algunos valores calculados los vemos en la siguiente tabla:

Tabla 3.2
Razón de la varianza con muestras panel y varianzas con muestras independientes
Estimador de promedio

Traslape	Coeficiente de Correlación			
	0.95	**0.85**	**0.75**	**0.7**
0.2	1.19	1.17	1.15	1.14
0.4	1.38	1.34	1.30	1.28
0.5	1.48	1.43	1.38	1.35
0.6	1.57	1.51	1.50	1.42
0.8	1.76	1.68	1.60	1.56
1	1.95	1.85	1.75	1.70

Figura 3.2: Razón de la varianza con muestras panel y varianzas con muestras independientes Estimador de promedio



En el caso del modelos panel 2 – (2) – 2 el incremento de la varianza del promedio de dos rondas es entre 35% y 48 % aproximadamente. En cambio el incremento con modelos panel 5 se encuentra entre 56% a 76% aproximadamente. Este incremento es más alto que en los modelos 2 – (2) – 2 porque el traslape entre rondas también es más alto.

Se podría realizar el mismo análisis en promedios de más rondas, como por ejemplo un estimador anual obtenido por medio del promedio de 4 trimestres.

3.2.2.1 El promedio de un año

Panel trimestral 2 – (2) – 2

En el caso de una encuesta panel trimestral 2 – (2) – 2 el estimador anual se obtiene mediante un promedio aritmético de los cuatro trimestres del año.

$$\hat{\theta}_{anual} = \frac{\hat{t}_1 + \hat{t}_2 + \hat{t}_3 + \hat{t}_4}{4} \qquad (3.13)$$

Este estimador tiene una varianza que se calcula de la siguiente manera:

$$Var\left[\hat{\theta}_{anual}\right] = Var\left[(\hat{t}_1 + \hat{t}_2 + \hat{t}_3 + \hat{t}_4)/4\right] = \frac{1}{16}\left(\sum_1^4 Var(\hat{t}_i) + 2\sum_{i<j} Cov(\hat{t}_1, \hat{t}_2)\right) \qquad (3.14)$$

Donde:

$$\sum_{i<j} Cov(\hat{t}_i, \hat{t}_j) = Cov(\hat{t}_1, \hat{t}_2) + Cov(\hat{t}_1, \hat{t}_3) + Cov(\hat{t}_1, \hat{t}_4) + Cov(\hat{t}_2, \hat{t}_3) + Cov(\hat{t}_2, \hat{t}_4) + Cov(\hat{t}_3, \hat{t}_4) \qquad (3.15)$$

Asumiendo igual varianza entre periodos teníamos la ecuación (3.3):

$$Cov(\hat{t}_{i+1},\hat{t}_i) = \beta_{i,i+1} \cdot \rho_{i,i+1} \cdot Var(\hat{t})$$

Reemplazando (3.3) en (3.15):

$$\sum_{i<j} Cov(\hat{t}_1,\hat{t}_2) = \beta_{12}\rho_{12}Var(\hat{t}) + \beta_{13}\rho_{13}Var(\hat{t}) + \beta_{14}\rho_{14}Var(\hat{t}) + \beta_{23}\rho_{23}Var(\hat{t}) + \\ + \beta_{24}\rho_{24}Var(\hat{t}) + \beta_{34}\rho_{34}Var(\hat{t})$$

$$\sum_{i<j} Cov(\hat{t}_1,\hat{t}_2) = Var(\hat{t})(\beta_{12}\rho_{12} + \beta_{13}\rho_{13} + \beta_{14}\rho_{14} + \beta_{23}\rho_{23} + \beta_{24}\rho_{24} + \beta_{34}\rho_{34}) \qquad (3.16)$$

donde:

β_{ij} : Porcentaje de traslape de la i-ésima con la j-ésima ronda

$\hat{\rho}_{ij}$: Coeficiente de correlación de la i-ésima con la j-ésima ronda

La ecuación (3.14) se reduce a lo siguiente:

$$Var\left[\hat{\theta}_{anual}\right] = \frac{1}{8}(2Var(\hat{t}) + Var(\hat{t})(\beta_{12}\rho_{12} + \beta_{13}\rho_{13} + \beta_{14}\rho_{14} + \beta_{23}\rho_{23} + \beta_{24}\rho_{24} + \beta_{34}\rho_{34})) \qquad (3.17)$$

Reemplazando los valores de traslape en el caso del panel trimestral 2 – (2) – 2, el estimador de la varianza del promedio es el siguiente:

$$Var\left[(\hat{t}_1 + \hat{t}_2 + \hat{t}_3 + \hat{t}_4)/4\right] = \frac{1}{16}\left(\sum_1^4 Var(\hat{t}_i) + 2\sum_{i<j} \mathrm{cov}(\hat{t}_1,\hat{t}_2)\right)$$

donde:

$\beta_{12} = 0.5$; $\beta_{23} = 0.5$; $\beta_{34} = 0.5$

$$\beta_{13}=0 \quad ; \quad \beta_{24}=0 \quad ; \quad \beta_{14}=0.25$$

Y además asumimos que:

$$Var(\hat{t}_i) \cong Var(\hat{t}) \qquad \forall i=1,2,3....$$

Entonces:

$$Var\left[(\hat{t}_1+\hat{t}_2+\hat{t}_3+\hat{t}_4)/4\right]=\frac{1}{16}(\sum_1^4 Var(\hat{t}_i)+2\sum_{i<j}\mathrm{cov}(\hat{t}_1,\hat{t}_2))$$

$$Var\left[(\hat{t}_1+\hat{t}_2+\hat{t}_3+\hat{t}_4)/4\right]=\frac{1}{16}(4Var(\hat{t})+2(Var(\hat{t})(\beta_{12}\rho_{12}+\beta_{13}\rho_{13}+\beta_{14}\rho_{14}+$$
$$+\beta_{23}\rho_{23}+\beta_{24}\rho_{24}+\beta_{34}\rho_{34}))$$

$$Var\left[(\hat{t}_1+\hat{t}_2+\hat{t}_3+\hat{t}_4)/4\right]=\frac{1}{8}Var(t)(2+0.5\rho_{12}+0.25\rho_{14}+0.5\rho_{23}+0.5\rho_{34})$$

$$Var\left[(\hat{t}_1+\hat{t}_2+\hat{t}_3+\hat{t}_4)/4\right]=\frac{1}{8}Var(\hat{t})(2+0.5(\rho_{12}+\rho_{23}+\rho_{34})+0.25\rho_{14}) \qquad (3.18)$$

De esta manera, la varianza para la estimación del promedio aumentará en relación directa con los coeficientes de correlación.

El promedio de un año para modelos con periodicidad mensual

Al igual que en los anteriores modelos de rotación con periodicidad trimestral se puede también calcular un indicador anual mediante un modelo de rotación que tenga periodicidad mensual. La diferencia será que el indicador calculado mediante el promedio será tomando en cuenta 12 rondas y no solamente 4 rondas, como era el anterior caso.

Entonces el estimador anual se obtiene por:

$$\hat{\theta}_{anual} = \frac{t_1 + t_2 + t_3 + t_4 + t_5 + t_6 + t_7 + t_8 + t_9 + t_{10} + t_{11} + t_{12}}{12} \Leftrightarrow n_i = n_j \forall i \neq j \tag{3.19}$$

Sin embargo sabemos que el año tiene un total de 52 semanas, este número no es múltiplo de 12, lo que quiere decir que algunos meses tendrán una semana más que otros meses. En varias encuestas con periodicidad trimestral se considera que cada tercer mes de cada trimestre del año tiene 5 semanas, y asumiendo un tamaño de muestra igual para todas las semanas, cuatro meses del año tienen mayor tamaño de muestra comparando con los meses que solamente tienen 4 semanas.

Entonces el estimador anual calculado mediante el promedio mensual tiene la siguiente forma:

$$\hat{\theta}_A = \frac{4}{52}\left(\hat{t}_i + \hat{t}_{i+1} + \hat{t}_{i+3} + \hat{t}_{i+4} + \hat{t}_{i+6} + \hat{t}_{i+7} + \hat{t}_{i+9} + \hat{t}_{i+10}\right) + \frac{5}{52}\left(\hat{t}_{i+2} + \hat{t}_{i+5} + \hat{t}_{i+8} + \hat{t}_{i+11}\right) \tag{3.20}$$

Donde la varianza es:

$$V(\hat{\theta})_A = V\left\{\frac{1}{52}\left[4\left(\hat{t}_i + \hat{t}_{i+1} + \hat{t}_{i+3} + \hat{t}_{i+4} + \hat{t}_{i+6} + \hat{t}_{i+7} + \hat{t}_{i+9} + \hat{t}_{i+10}\right) + 5\left(\hat{t}_{i+2} + \hat{t}_{i+5} + \hat{t}_{i+8} + \hat{t}_{i+11}\right)\right]\right\}$$

$$V(\hat{\theta})_A = \frac{1}{52^2}\left[\begin{array}{l} 16 \cdot V\left(\hat{t}_i + \hat{t}_{i+1} + \hat{t}_{i+3} + \hat{t}_{i+4} + \hat{t}_{i+6} + \hat{t}_{i+7} + \hat{t}_{i+9} + \hat{t}_{i+10}\right) + 25 \cdot V\left(\hat{t}_{i+2} + \hat{t}_{i+5} + \hat{t}_{i+8} + \hat{t}_{i+11}\right) \\ + 2Cov(\hat{t}_i + \hat{t}_{i+1} + \hat{t}_{i+3} + \hat{t}_{i+4} + \hat{t}_{i+6} + \hat{t}_{i+7} + \hat{t}_{i+9} + \hat{t}_{i+10}, \hat{t}_{i+2} + \hat{t}_{i+5} + \hat{t}_{i+8} + \hat{t}_{i+11}) \end{array}\right]$$

Si suponemos que la varianza es igual para cada ronda

$$\forall i = 1,2,3.... \qquad Var(\hat{t}_i) = Var(\hat{t})$$

Entonces:

$$V(\hat{\theta})_A = \frac{1}{52^2}\left[16 \cdot 8 \cdot V\left(\hat{t}\right) + 25 \cdot 4 \cdot V\left(\hat{t}\right) + 2(\sum_{i<j} Cov(\hat{t}_i, \hat{t}_j)\right]$$

$$V(\hat{\theta})_A = \frac{1}{52^2}\left[228 \cdot V\left(\hat{t}\right) + 2\left(\sum\nolimits_{i<j} Cov(\hat{t}_i, \hat{t}_j)\right)\right] \tag{3.21}$$

Bajo el supuesto:

$$Cov(\hat{t}_{j+1}, \hat{t}_i) = \beta_{i,i+1} \cdot \rho_{i,i+1} \cdot Var(\hat{t}) = \beta_1 \cdot \rho_1 \cdot Var(\hat{t}) \qquad \forall i$$

$$Cov(\hat{t}_{i+2}, \hat{t}_i) = \beta_{i,i+2} \cdot \rho_{i,i+2} \cdot Var(\hat{t}) = \beta_2 \cdot \rho_2 \cdot Var(\hat{t}) \qquad \forall i$$

.
.
.

$$Cov(\hat{t}_{i+n}, \hat{t}_i) = \beta_{i,i+n} \cdot \rho_{i,i+n} \cdot Var(\hat{t}) = \beta_n \cdot \rho_n \cdot Var(\hat{t}) \qquad \forall i, n = 1,2,...,11$$

(3.22)

Esto quiere decir que se asume que la covarianza entre la ronda 1 y la ronda 2 es igual que la covarianza entre la ronda 4 y la ronda 5, o que la covarianza entre la ronda 5 y la ronda 9 es igual que la covarianza entre la ronda 3 y la ronda 7. En general, la covarianza entre rondas separadas en igual intervalo de tiempo es constante.

Reemplazando (3.22) en (3.21) se tiene:

$$V(\hat{\theta})_A = \frac{1}{52^2}\left[228 \cdot V\left(\hat{t}\right) + 2\left(\sum_{n=1}^{11}(12-n)\beta_n \cdot \rho_n \cdot Var(\hat{t})\right)\right]$$

$$V(\hat{\theta})_A = \frac{1}{52^2}\left[228 \cdot V\left(\hat{t}\right) + 2 \cdot V(\hat{t})\left(\sum_{n=1}^{11}(12-n)\beta_n \cdot \rho_n\right)\right]$$

$$V(\hat{\theta})_A = \frac{57 \cdot V\left(\hat{t}\right)}{676}\left[1 + \frac{1}{144}\left(\sum_{n=1}^{11}(12-n)\beta_n \cdot \rho_n\right)\right] \tag{3.23}$$

De esta forma se realiza el cálculo de la varianza del promedio.

Veamos el caso para un modelo panel mensual 4 – (8) – 4, la varianza tendrá la siguiente forma:

$$V(\hat{\theta})_A = \frac{57 \cdot V(\hat{t})}{676}\left[1 + \frac{1}{144}\left(\sum_{n=1}^{11}(12-n)\beta_n \cdot \rho_n\right)\right]$$

$$V(\hat{\theta})_A = \frac{57 \cdot V(\hat{t})}{676}\left[\begin{array}{l} 1 + \frac{1}{144}(11\beta_1\rho_1 + 10\beta_2\rho_2 + 9\beta_3\rho_3 + 8\beta_4\rho_4 + 7\beta_5\rho_5 + 6\beta_6\rho_6 + \\ + 5\beta_7\rho_7 + 4\beta_8\rho_8 + 3\beta_9\rho_9 + 2\beta_{10}\rho_{10} + \beta_{11}\rho_{11}) \end{array}\right]$$

Según el modelo de rotación no existe la covarianza entre periodos con distancia de 4, 5, 6, 7 y 8 rondas, para los cuales el porcentaje de traslape para es $\beta_n = 0$ *donde* n = *4,5,6,7,8*

El traslape entre rondas es el siguiente:

$\beta_1 = 0.75$	$\beta_2 = 0.50$	$\beta_3 = 0.25$
$\beta_9 = 0.125$	$\beta_{10} = 0.25$	$\beta_{11} = 0.375$

Reemplazando

$$V(\hat{\theta})_A = \frac{57 \cdot V(\hat{t})}{676}\left[1 + \frac{1}{144}(8.25\rho_1 + 5\rho_2 + 2.25\rho_3 + 0.375\rho_9 + 0.5\rho_{10} + 0.375\rho_{11})\right]$$

$$V(\hat{\theta})_A = \frac{57 \cdot V(\hat{t})}{676}\left[1 + \frac{1}{144}(8.25\rho_1 + 5\rho_2 + 2.25\rho_3 + 0.375(\rho_9 + \rho_{11}) + 0.5\rho_{10}\right] \qquad (3.24)$$

Lo que resta por hacer es reemplazar los valores calculados con los datos obtenidos.

3.2.3 Estimador combinado

En varios de los modelos de rotación podemos ver que en rondas consecutivas existe un porcentaje de traslape de la muestra, es decir, un porcentaje de la muestra participa en la ronda *i* y en ronda *i+1*. Por tal efecto, podemos aprovechar este hecho para mejorar el estimador de una ronda utilizando información de una ronda anterior.
En esta sección veremos el estimador combinado[6] para la media.

Notación
Para esta sección usaremos las siguientes notaciones:

n : Tamaño de la muestra en cada ronda

n_i : Número de elementos que se reemplazan en la ronda i respecto a la ronda i-1

m_i : Número de elementos que se conservan en la ronda i respecto a la ronda i-1

β_i : Proporción de muestra en común de la ronda *i* respecto a la ronda *i-1*

ρ_i : Coeficiente de correlación entre la ronda *i* e *i-1*

En este entendido se tiene lo siguiente:

$$m_i = n\beta_i \qquad ; \qquad 0 < \beta_i < 1$$

$$n_i = n \cdot (1 - \beta_i) \qquad ; \qquad 0 < \beta_i < 1$$

El estimador combinado propuesto es el siguiente:

[6] Este estimador fue planteado en la tesis de Angulo A. (1999)

$$\hat{\hat{t}}_{C,i+1} = \phi \cdot \hat{\hat{t}}_{i+1}^{n_1} + (1-\phi)\hat{\hat{t}}_{i+1}^{m} \qquad 0 \le \phi \le 1 \qquad (3.25)$$

Con varianza igual a :

$$V(\hat{\hat{t}}_{C,i+1}) = \phi^2 \cdot V(\hat{\hat{t}}_{i+1}^{n_1}) + (1-\phi)^2 V(\hat{\hat{t}}_{i+1}^{m}) \qquad (3.26)$$

donde

- $\hat{\hat{t}}_{i+1}^{n_1}$: Estimador del promedio de la (i+1) ésima ronda de la parte no traslapada con la ronda i

$$\hat{t}_{i+1}^{n_1} = \sum_{j=1}^{n_i} \frac{t_{j,i+1}}{n_i}$$

y varianza:

$$V(\hat{t}_{i+1}^{n_1}) = \frac{\sigma^2}{n\alpha_i} \qquad (3.27)$$

donde: $\alpha_i = 1 - \beta_i$

- $\hat{\hat{t}}_{i+1}^{m}$: Estimador de la (*i+1*) ésima ronda basada en la diferencia de la parte común y en la primera muestra seleccionada en la primera ocasión

$$\hat{\hat{t}}_{i+1}^{m} = \sum_{j=1}^{n} \frac{t_{j,i}}{n} + \left(\frac{1}{m_i}\sum_{j=1}^{m_i} t_{j,i+1} - \frac{1}{m_i}\sum_{j=1}^{m_i} t_{j,i}\right) \qquad (3.28)$$

donde

$\sum_{j=1}^{n} \frac{t_{j,i}}{n}$:	Promedio de la ronda i
$\frac{1}{m_i}\sum_{j=1}^{m_i} t_{j,i+1}$:	Promedio de la ronda i+1 de la parte traslapada

$\frac{1}{m_i}\sum_{j=1}^{m_i} t_{j,i}$: Promedio de la ronda i de la parte traslapada

y varianza

$$V(\hat{\hat{t}}_{i+1}^{m}) = \frac{\sigma^2}{n\beta_i}\left[1+\alpha_i(1-2\rho_i)\right] \tag{3.27}$$

Para demostrar que la ecuación (3.27) es la varianza del estimador combinado primeramente se demuestra que el estimador $\hat{\hat{t}}_{i+1}^{m}$ es un estimador insesgado.

Teorema 3.1

El valor esperado de cualquier variable aleatoria X se puede calcular:

$$E(X) = E(E(X/u_i) = E_1E_2(X) \tag{3.28}$$

La demostración de este teorema está en el anexo 3.1

Entonces primeramente se calcula el valor esperado de $\hat{\hat{t}}_{i+1}^{m}$ dado el evento $\hat{\hat{t}}_{i}$

$$E_2(\hat{\hat{t}}_{i+1}^{m}) = E\left[\sum_{j=1}^{n}\frac{t_{j,i}}{n} + (\frac{1}{m_i}\sum_{j=1}^{m_i} t_{j,i+1} - \frac{1}{m_i}\sum_{j=1}^{m_i} t_{j,i})\right]$$

$$E_2(\hat{\hat{t}}_{i+1}^{m}) = E(\sum_{j=1}^{n}\frac{t_{j,i}}{n}) + E(\frac{1}{m_i}\sum_{j=1}^{m_i} t_{j,i+1}) - E(\frac{1}{m_i}\sum_{j=1}^{m_i} t_{j,i})$$

$$E_2(\hat{\hat{t}}_{i+1}^{m}) = \sum_{j=1}^{n}\frac{t_{j,i}}{n} + \frac{1}{n}\sum_{j=1}^{n_i} t_{j,i+1} - \frac{1}{n}\sum_{j=1}^{n} t_{j,i}$$

$$E_2(\hat{\hat{t}}_{i+1}^{m}) = \frac{1}{n}\sum_{j=1}^{n_i} t_{j,i+1} \tag{3.29}$$

Extrayendo el segundo valor esperado tenemos:

$$E_1 E_2(\hat{\bar{t}}_{i+1}^m) = E(\frac{1}{n}\sum_{j=1}^{n_i} t_{j,i+1}) = \bar{t}_{i+1} \tag{3.30}$$

Lo que muestra que el estimador $\hat{\bar{t}}_{i+1}^m$ es un estimador insesgado de $\bar{t}_{i+1}$.

La varianza la calcularemos usando el siguiente teorema:

Teorema 3.2

$$\mathrm{cov}(X,Y) = E_1[\mathrm{cov}_2(X,Y)] + \mathrm{cov}_1[E_2(X), E_2(Y)]$$

Para la demostración del teorema 3.2 ver Anexo 3.2

Aplicando el corolario del teorema 3.2 (ver Anexo 3.2) tenemos lo siguiente:

$$V(\hat{\bar{t}}_{i+1}^m) = E_1 V_2(\hat{\bar{t}}_{i+1}^m) + V_1 E_2(\hat{\bar{t}}_{i+1}^m) \tag{3.31}$$

donde $$V_1 E_2(\hat{\bar{t}}_{i+1}^m) = V_1(\frac{1}{n}\sum_{j=1}^{n_i} t_{j,i+1}) = \frac{\sigma^2}{n} \tag{3.32}$$

Ahora queda calcular $E_1 V_2(\hat{\bar{t}}_{i+1}^m)$ de la siguiente forma:

$$V_2(\hat{\bar{t}}_{i+1}^m) = (\frac{1}{m_i} - \frac{1}{n})\frac{1}{n-1}\sum_{j=1}^{n}\left[t_{j,i+1} - t_{j,i} - \frac{1}{n}\sum_{k=1}^{n}(t_{j,i+1} - t_{j,i})\right]^2 \qquad (3.33)$$

Calculando el valor esperado se tiene:

$$E_1V_2(\hat{\bar{t}}_{i+1}^m) = (\frac{1}{m_i} - \frac{1}{n})\frac{1}{N-1}\sum_{j=1}^{N}\left[(t_{j,i+1} - \bar{t}_{i+1}) - (t_{j,i} - \bar{t}_i)\right]^2$$

$$E_1V_2(\hat{\bar{t}}_{i+1}^m) = (\frac{1}{m_i} - \frac{1}{n})(2\sigma^2 - 2\rho_i\sigma^2)$$

$$E_1V_2(\hat{\bar{t}}_{i+1}^m) = \frac{n - m_i}{nm_i}2\sigma^2(1-\rho_i) = \frac{2\sigma^2}{n\beta_i}(1-\beta_i)(1-\rho_i)$$

(3.34)

Reemplazando los valores de la ecuación (3.32) y (3.34) en la ecuación (3.31):

$$V(\hat{\bar{t}}_{i+1}^m) = \frac{2\sigma^2}{n\beta_i}(1-\beta_i)(1-\rho_i) + \frac{\sigma^2}{n} = \frac{\sigma^2}{n}\left[\frac{\beta_i + 2(1-\beta_i)(1-\rho_i)}{\beta_i}\right]$$

$$V(\hat{\bar{t}}_{i+1}^m) = \frac{\sigma^2}{n\beta_i}(1 + (1-\beta_i)(1-2\rho_i)) \qquad \text{Si} \quad \alpha_i = 1 - \beta_1$$

$$V(\hat{\bar{t}}_{i+1}^m) = \frac{\sigma^2}{n\beta_i}\left[1 + \alpha_i(1-2\rho_i)\right]$$

Queda demostrada la varianza de $\hat{\bar{t}}_{i+1}^m$

Una vez definido cada uno de los términos del estimador combinado, en el siguiente paso se calcula el valor de ϕ que minimiza la varianza. Esto se logra derivando (3.26) respecto de ϕ e igualando a acero.

$$\frac{dV(\hat{\hat{t}}_{C,i+1})}{d\phi} = 0$$

$$2 \cdot \phi \cdot V(\hat{\hat{t}}^{n_1}_{i+1}) - 2(1-\phi)V(\hat{\hat{t}}^{m}_{i+1}) = 0$$

$$2 \cdot \phi \cdot V(\hat{t}^{n_1}_{i+1}) - 2V(\hat{t}^{m}_{i+1}) + 2\phi \cdot V(\hat{t}^{m}_{i+1}) = 0 \qquad //1/2$$

$$\phi \cdot V(\hat{t}^{n_1}_{i+1}) + \phi \cdot V(\hat{t}^{m}_{i+1}) = V(\hat{t}^{m}_{i+1})$$

$$\phi = \frac{V(\hat{\hat{t}}^{m}_{i+1})}{V(\hat{\hat{t}}^{n_1}_{i+1}) + V(\hat{\hat{t}}^{m}_{i+1})} \tag{3.35}$$

Calculamos la segunda derivada:

$$\frac{d^2V(\hat{\hat{t}}_{C,i+1})}{d^2\phi} = 2 \cdot V(\hat{\hat{t}}^{n_1}_{i+1}) + 2V(\hat{\hat{t}}^{m}_{i+1}) > 0$$

Con esto comprobamos que el valor calculado para ϕ es un valor que minimiza la varianza del estimador combinado.

Reemplazando ϕ que minimiza la varianza en la ecuación (3.26) se tiene:

$$V(\hat{\hat{t}}_{C,i+1}) = \left(\frac{V(\hat{\hat{t}}_{i+1}^{m})}{V(\hat{\hat{t}}_{i+1}^{n_1}) + V(\hat{\hat{t}}_{i+1}^{m})}\right)^2 \cdot V(\hat{\hat{t}}_{i+1}^{n_1}) + \left(1 - \frac{V(\hat{\hat{t}}_{i+1}^{m})}{V(\hat{\hat{t}}_{i+1}^{n_1}) + V(\hat{\hat{t}}_{i+1}^{m})}\right)^2 V(\hat{\hat{t}}_{i+1}^{m})$$

$$V(\hat{\hat{t}}_{C,i+1}) = \frac{V^2(\hat{\hat{t}}_{i+1}^{m})}{(V(\hat{\hat{t}}_{i+1}^{n_1}) + V(\hat{\hat{t}}_{i+1}^{m}))^2} \cdot V(\hat{\hat{t}}_{i+1}^{n_1}) + \frac{V^2(\hat{\hat{t}}_{i+1}^{n})}{(V(\hat{\hat{t}}_{i+1}^{n_1}) + V(\hat{\hat{t}}_{i+1}^{m}))^2} V(\hat{\hat{t}}_{i+1}^{m})$$

$$V(\hat{\hat{t}}_{C,i+1}) = \frac{V^2(\hat{\hat{t}}_{i+1}^{m}) \cdot V(\hat{\hat{t}}_{i+1}^{n_1}) + V(\hat{\hat{t}}_{i+1}^{m}) \cdot V^2(\hat{\hat{t}}_{i+1}^{n_1})}{(V(\hat{\hat{t}}_{i+1}^{n_1}) + V(\hat{\hat{t}}_{i+1}^{m}))^2}$$

$$V(\hat{\hat{t}}_{C,i+1}) = \frac{V(\hat{\hat{t}}_{i+1}^{m}) \cdot V(\hat{\hat{t}}_{i+1}^{n_1})(V(\hat{\hat{t}}_{i+1}^{m}) + V(\hat{\hat{t}}_{i+1}^{n_1}))}{(V(\hat{\hat{t}}_{i+1}^{n_1}) + V(\hat{\hat{t}}_{i+1}^{m}))^2}$$

$$V(\hat{\hat{t}}_{C,i+1}) = \frac{V(\hat{\hat{t}}_{i+1}^{m}) \cdot V(\hat{\hat{t}}_{i+1}^{n_1})}{V(\hat{\hat{t}}_{i+1}^{n_1}) + V(\hat{\hat{t}}_{i+1}^{m})}$$

Reemplazando los valores de la varianza de $\hat{\hat{t}}_{i+1}^{m}$ y $\hat{\hat{t}}_{i+1}^{n_1}$ tenemos:

$$V(\hat{\hat{t}}_{C,i+1}) = \frac{\frac{\sigma^2}{n\beta_i}\left[1 + \alpha_i(1 - 2\rho_i)\right]\frac{\sigma^2}{n\alpha_i}}{\frac{\sigma^2}{n\alpha_i} + \frac{\sigma^2}{n\beta_i}\left[1 + \alpha_i(1 - 2\rho_i)\right]} = \frac{\frac{\sigma^2}{n\beta_i}\left[1 + \alpha_i(1 - 2\rho_i)\right]\frac{1}{\alpha_i}}{\frac{1}{\alpha_i} + \frac{1}{\beta_i}\left[1 + \alpha_i(1 - 2\rho_i)\right]}$$

$$V(\hat{\hat{t}}_{C,i+1}) = \frac{\frac{\sigma^2}{n\beta_i\alpha_i} + \frac{\sigma^2}{n\beta_i}(1 - 2\rho_i)}{\frac{1}{\alpha_i} + \frac{1}{\beta_i}\left[1 + \alpha_i(1 - 2\rho_i)\right]} = \frac{\frac{\sigma^2 + \sigma^2\alpha_i(1 - 2\rho_i)}{n\beta_i\alpha_i}}{\frac{\beta_i + \alpha_i\left[1 + \alpha_i(1 - 2\rho_i)\right]}{\beta_i\alpha_i}}$$

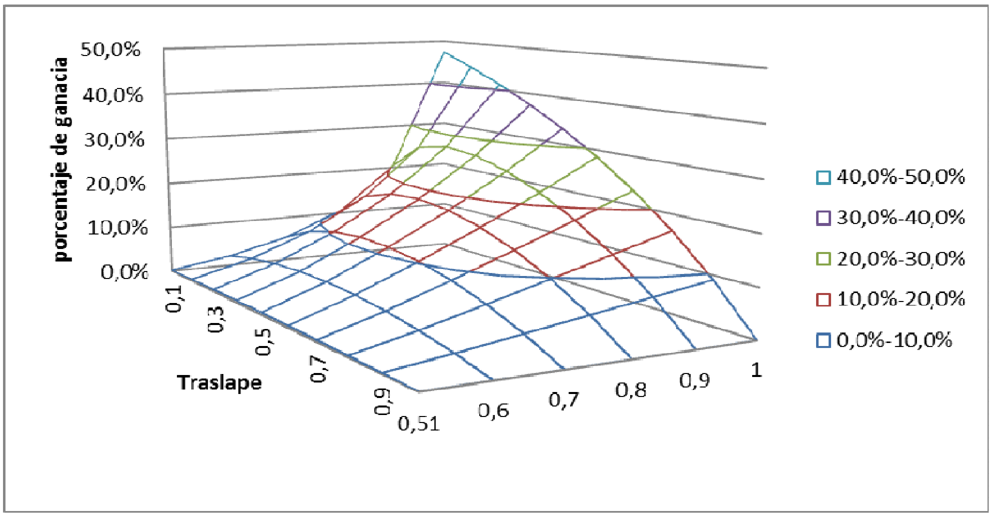
$$V(\hat{\hat{t}}_{C,i+1}) = \frac{\sigma^2 + \sigma^2\alpha_i(1 - 2\rho_i)}{n \cdot \left\{\beta_i + \alpha_i\left[1 + \alpha_i(1 - 2\rho_i)\right]\right\}} = \frac{\sigma^2(1 + \alpha_i(1 - 2\rho_i))}{n\beta_i + n\alpha_i\left[1 + \alpha_i(1 - 2\rho_i)\right]}$$

$$V(\hat{\bar{t}}_{C,i+1}) = \frac{\sigma^2(1+\alpha_i(1-2\rho_i))}{n(\beta_i+\alpha_i[1+\alpha_i(1-2\rho_i)])} \tag{3.36}$$

Entonces la varianza del estimador combinado para la segunda ocasión es como muestra la ecuación (3.36).

Nótese que si β_i =1 (traslape total), entonces la varianza del estimador combinado es igual a σ^2/n. Ahora si tomamos dos muestras independientes para cada ronda (esto quiere decir que β_i = 0), la varianza toma el mismo valor de σ^2/n. Entonces para uno o para otro caso la precisión del estimador no mejora ni empeora.
Veamos gráficamente:

Figura 3.3: Ganancia en precisión



3.2.3.1 Porcentaje Optimo de Traslape

Notamos que el estimador combinado obtiene la misma precisión con muestras independientes que con muestras totalmente traslapadas. Entonces corresponde calcular cuál es el porcentaje de traslape óptimo entre rondas de tal forma que la varianza del estimador combinado se haga mínima.

Diferenciamos la ecuación (3.36) respecto de α_i e igualando a cero se tiene:

$$\frac{dV(\hat{\hat{t}}_{C,i+1})}{d\alpha_i} = 0$$

Donde:

$$V(\hat{\hat{t}}_{C,i+1}) = \frac{\sigma^2(1+\alpha_i(1-2\rho_i))}{n(1-\alpha_i)+n\alpha_i\left[1+\alpha_i(1-2\rho_i)\right]}$$

$$(1-2\rho_i)\left[1+\alpha_i^2(1-2\rho_i)\right]-2\alpha_i(1-2\rho_i)\left[1+\alpha_i(1-2\rho_i)\right]=0$$

$$(1-2\rho_i)+\alpha_i^2(1-2\rho_i)^2-2\alpha(1-2\rho_i)-2\alpha_i^2(1-2\rho_i)^2=0$$

$$(1-2\rho_i)-\alpha_i^2(1-2\rho_i)^2(-1)-2\alpha_i(1-2\rho_i)=0 \qquad //(-1)$$

$$\alpha_i^2(1-2\rho_i)^2+2\alpha_i(1-2\rho_i)-(1-2\rho_i)=0$$

Resolviendo la ecuación:

$$\alpha_i = \frac{\sqrt{2-2\rho_i}-1}{1-2\rho_i} \qquad \text{con} \qquad \rho_i \neq 0.5$$

Por lo tanto el traslape óptimo que minimiza la varianza es:

$$\beta_i = \frac{2(1-\rho_i)-\sqrt{2-2\rho_i}}{1-2\rho_i} \qquad \text{con} \qquad \rho_i \neq 0.5 \qquad (3.37)$$

Reemplazando en la ecuación (3.30) para calcular la varianza mínima se tiene:

$$V(\hat{\hat{t}}_{C,i+1}) = \frac{\sigma^2(1+\alpha_i(1-2\rho_i))}{n\beta_i + n\alpha_i\left[1+\alpha_i(1-2\rho_i)\right]}$$

$$V(\hat{\hat{t}}_{C,i+1}) = \frac{\sigma^2}{n}\cdot\frac{1+\frac{\sqrt{2-2\rho_i}-1}{1-2\rho_i}\cdot(1-2\rho_i)}{\frac{2(1-\rho_i)-\sqrt{2-2\rho_i}}{1-2\rho_i}+\frac{\sqrt{2-2\rho_i}-1}{1-2\rho_i}(1+\frac{\sqrt{2-2\rho_i}-1}{1-2\rho_i}\cdot(1-2\rho_i)}$$

$$V(\hat{\hat{t}}_{C,i+1}) = \frac{\sigma^2}{n}\cdot\frac{1+\frac{\sqrt{2-2\rho_i}-1}{1-2\rho_i}\cdot(1-2\rho_i)}{1+\left(\frac{\sqrt{2-2\rho_i}-1}{1-2\rho_i}\right)^2\cdot(1-2\rho)} = \frac{\sigma^2}{n}\cdot\frac{1+\sqrt{2-2\rho_i}-1}{1+\frac{(\sqrt{2-2\rho_i}-1)^2}{1-2\rho_i}}$$

$$V(\hat{\hat{t}}_{C,i+1}) = \frac{\sigma^2}{n}\cdot\frac{(1-2\rho_i)\sqrt{2-2\rho_i}}{2(2-2\rho_i)-2\sqrt{2-2\rho_i}} = \frac{\sigma^2}{n}\cdot\frac{(1-2\rho_i)\sqrt{2-2\rho_i}}{2\sqrt{2-2\rho_i}(\sqrt{2-2\rho_i}-1)}$$

$$V(\hat{\hat{t}}_{C,i+1}) = \frac{\sigma^2}{n}\cdot\frac{(1-2\rho_i)}{2(\sqrt{2-2\rho_i}-1)} \qquad (3.38)$$

La varianza del estimador combinado será más pequeña que de una muestra independiente si ρ_i >0.5.

3.2.3.2 Ganancia en precisión

La ganancia en precisión G del estimador combinado $\hat{\hat{t}}_{C,i+1}$, que utiliza información de una ronda anterior y además tiene traslape óptimo entre rondas, sobre el estimador simple $\hat{t}_{MAS(i+1)}$ de muestras independientes, se calcula mediante la siguiente expresión:

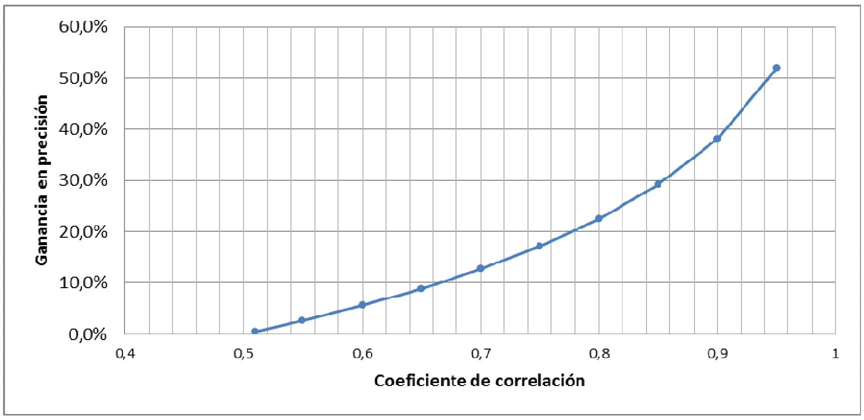
$$G = \frac{V(\hat{\hat{t}}_{MAS(i+1)}) - V(\hat{\hat{t}}_{C,i+1})}{V(\hat{\hat{t}}_{C,i+1})} = \frac{2(\sqrt{2-2\rho_i} + \rho_i) - 3}{1 - 2\rho_i} \tag{3.39}$$

Algunos valores posibles son:

Tabla 3.3
Ganancia en precisión (en porcentaje)

Coeficiente de correlación	% Traslape óptimo	% ganancia en precisión G
0.51	49.7%	0.5%
0.55	48.7%	2.6%
0.6	47.2%	5.6%
0.65	45.6%	8.9%
0.7	43.6%	12.7%
0.75	41.4%	17.2%
0.8	38.7%	22.5%
0.85	35.4%	29.2%
0.9	30.9%	38.2%
0.95	24.0%	51.9%

Figura 3.4: Ganancia en precisión con traslape óptimo (porcentaje)



Para valores bajos de correlación la ganancia en precisión es también baja, sin embargo para valores altos de correlación y tomando en cuenta el traslape óptimo, la ganancia en precisión es mayor.

Capítulo 4

Aplicación

4.1 Introducción

La Encuesta de Empleo (anteriormente llamada Encuesta Trimestral de Empleo) se fue realizando desde enero del 2009 hasta julio 2011 por el Instituto Nacional de Estadística. Esta encuesta tiene la característica de ser tipo panel rotatorio con traslape parcial, justamente lo que se necesita para lograr una aplicación de todo lo desarrollado a lo largo del documento. Las bases de datos proporcionadas por el Instituto Nacional de Estadística corresponden a los dos primeros trimestres del año 2010 de la Encuesta Trimestral de Empleo.

4.2 Características de la Encuesta

Según el documento muestral colgado en la página del INE[7], la encuesta de Empleo cumple con toda la rigurosidad de un diseño muestral que es: Un marco muestral adecuado, cálculo del tamaño de muestra, tipo de muestreo, construcción de los estimadores, etc.

[7] http://www.ine.gob.bo:8081/etePortal/documentacion/disenomuestral.pdf

4.2.1 Tipo de Muestreo

El muestreo utilizado en la Encuesta Trimestral de Empleo es un muestreo bietápico, estratificado y por conglomerados[8].

4.2.2 Unidades de muestreo y unidades de análisis

Las unidades de muestreo en la primera etapa son los conglomerados de manzanos llamadas unidades primarias de muestreo (UPM), en la segunda etapa de selección las unidades de muestreo son las viviendas. Las unidades de análisis son las personas mayores a 14 años que habitan dentro de la vivienda. Sin embargo se registra a todas las personas que habitan dentro de la vivienda.

Para la aplicación de la parte de los estimadores no se tomará en cuenta el diseño muestral de la encuesta y se trabajará bajo el supuesto de un muestreo aleatorio simple de viviendas.

4.2.3 Periodicidad de la encuesta

Desde el inicio de la encuesta hasta el cuarto trimestre del 2010, la encuesta generaba resultados de manera trimestral. Posteriormente, desde el primer trimestre 2011, la encuesta además de reportar datos trimestrales también genera resultados mensuales sin haber alterado el modelo de rotación.

Según el diseño muestral de la Encuesta de Empleo, la recolección de información se realizaba de manera continua, donde cada trimestre estaba conformado por 12 semanas, tomando la última semana como parada técnica. A partir del primer

---

[8] Para mayor detalle consultar la página
http://www.ine.gob.bo:8081/etePortal/documentacion/disenomuestral.pdf

trimestre de 2011 se recolecta información de las 13 semanas de cada trimestre, anulando así la parada técnica y cubriendo totalmente las 52 semanas del año.

4.2.4 Modelo de Rotación de la Encuesta de Empleo

El Modelo de rotación adoptado en la Encuesta de Empleo es el modelo 2 – (2) – 2. Esto implica que cada unidad muestral se visite un total de cuatro trimestres; dos trimestres consecutivos, después descansa los dos trimestres posteriores, para finalmente ingresar en la encuesta los siguientes dos trimestres consecutivos.

El modelo de rotación es el siguiente:

Tabla 4.1
Encuesta Trimestral de Empleo – Modelo de Rotación

PERIODO	VISITA			
	1º	**2º**	**3º**	**4º**
	PANEL			
1º - 2009	P11	P12	P13	P14
2º - 2009	P15	P11	P16	P13
3º - 2009	P17	P15	P18	P16
4º - 2009	P19	P17	P12	P18
1º - 2010	**P20**	P19	**P11**	P12
2º - 2010	P21	**P20**	P15	**P11**
3º - 2010	P22	P21	P17	P15
4º - 2010	P23	P22	P19	P17
1º - 2011	P24	P23	**P20**	P19
2º - 2011	P25	P24	P21	**P20**
3º - 2011	P26	P25	P22	P21
4º - 2011	P27	P26	P23	P22
1º - 2012	P28	P27	P24	P23
2º - 2012	P29	P28	P25	P24
3º - 2012	P30	P29	P26	P25
4º - 2012	P31	P30	P27	P26

Los trimestres en el círculo son los trimestres con sus respectivos paneles donde se realizará la aplicación.

4.2.5 Balanceado y distribución de la muestra

Una vez seleccionada la muestra, según el documento del diseño muestral, la muestra es distribuida de manera aleatoria en los paneles: P20, P19, P11 y P12 en el primer trimestre, y los paneles P21, P20, P15 y P11 del segundo trimestre. Además la distribución en las 12 semanas de cada trimestre también es aleatoria. Como la base proporcionada es del primer y segundo trimestre 2010, la distribución de la muestra se realizó solamente en 12 semanas.

Por aspectos de confidencialidad no se cuenta con la base completa. La base proporcionada corresponde a la ciudad de La Paz. Esta pequeña muestra fue seleccionada de la base de datos de la ciudad de la Paz. Por tal motivo esta pequeña muestra proporcionada por el Instituto Nacional de Estadística no está bien balanceada, porque al ser una muestra pequeña que se extrajo de manera aleatoria, no se consideró ningún tipo de control con respecto al balanceado por panel ni semana de referencia.

El balanceado de la muestra total es como muestra la Tabla 4.2 y la Tabla 4.3. Se puede observar que la muestra está bien balanceada en cuanto a unidades primarias de muestreo. Según el diseño muestral[9], en la segunda etapa se selecciona 9 viviendas de cada unidad primaria y se encuesta a todas las personas que habitan dentro la vivienda. Es claro que al igual que el balanceado de la muestra por unidades primarias de muestreo, la muestra en términos de viviendas está correctamente balanceada

[9] http://www.ine.gob.bo:8081/etePortal/documentacion/disenomuestral.pdf

Tabla 4.2
Distribución de las unidades primarias de muestreo de la muestra 1er trimestre 2010 ciudad La Paz

		PANEL				
		20	**19**	**11**	**12**	**Total**
SEMANA	**1**	5	5	5	5	20
	2	5	5	5	5	20
	3	5	5	5	5	20
	4	5	5	5	5	20
	5	5	5	5	5	20
	6	5	5	5	5	20
	7	5	5	5	5	20
	8	5	5	5	5	20
	9	5	5	5	5	20
	10	5	5	5	5	20
	11	5	5	5	5	20
	12	5	5	5	5	20
Total		**60**	**60**	**60**	**60**	**240**

Fuente: Encuesta de Trimestral de Empleo 1/2010
Instituto Nacional de Estadística

Tabla 4.3
Distribución de las unidades primarias de muestreo de la muestra 2do trimestre 2010 ciudad La Paz

		PANEL				
		20	**19**	**11**	**12**	Total
SEMANA	**1**	5	5	5	5	20
	2	5	5	5	5	20
	3	5	5	5	5	20
	4	5	5	5	5	20
	5	5	5	5	5	20
	6	5	5	5	5	20
	7	5	5	5	5	20
	8	5	5	5	5	20
	9	5	5	5	5	20
	10	5	5	5	5	20
	11	5	5	5	5	20
	12	5	5	5	5	20
Total		**60**	**60**	**60**	**60**	**240**

Fuente: Encuesta de Trimestral de Empleo 2/2010
Instituto Nacional de Estadística

Para la aplicación con los datos del INE, se cuenta con 50 unidades primarias de muestreo en cada trimestre, haciendo un total de 394 viviendas para cada uno de los trimestres.

Veamos ahora la distribución de muestra que proporcionó el INE para la aplicación del presente documento.

Tabla 4.4
Distribución de las unidades primarias de muestreo de la muestra proporcionada del 1er trimestre 2010

		PANEL				
		11	**12**	**19**	**20**	**Total**
SEMANA	**1**	0	1	0	3	4
	2	2	1	3	0	6
	3	2	1	3	1	7
	4	1	1	1	2	5
	5	3	2	1	0	6
	6	1	0	1	1	3
	7	1	0	1	0	2
	8	1	2	1	2	6
	9	1	1	0	1	3
	10	0	1	1	0	2
	11	1	0	0	0	1
	12	1	1	1	2	5
Total		**14**	**11**	**13**	**12**	**50**

Fuente: Encuesta de Trimestral de Empleo 2/2010
Instituto Nacional de Estadística

Tabla 4.5
Distribución de las unidades primarias de muestreo de la muestra proporcionada del 2do trimestre 2010

		PANEL				
		11	**15**	**20**	**21**	Total
SEMANA	**1**	1	0	3	0	4
	2	2	1	0	0	3
	3	4	1	1	1	7
	4	1	0	2	0	3
	5	3	2	2	1	8
	6	1	1	1	0	3
	7	1	1	1	0	3
	8	1	2	2	0	5
	9	1	1	1	0	3
	10	1	1	1	0	3
	11	2	2	0	0	4
	12	1	1	2	0	4
Total		**19**	**13**	**16**	**2**	**50**

Fuente: Encuesta de Trimestral de Empleo
Instituto Nacional de Estadística

Es claro que el balanceado de la muestra no está uniforme por panel ni por semana de referencia. De seguro esto ocurre porque la muestra solicitada fue seleccionada de manera aleatoria. Para fines de aplicación esto no interesa, porque tiene que quedar claro que el correcto balanceo se realizó en la muestra total, y eso lo comprobamos en la Tablas 4.2 y la Tabla 4.3.

Aplicando las reglas de balanceado de la muestra propuestas en el capítulo 2, la muestra correctamente balanceada debería tener la distribución que se aprecia en la Tabla 4.6. Sin embargo, como se mencionó en un párrafo anterior, la muestra que se seleccionó de la base de datos de la ETE, se la realizó de manera aleatoria, sin ningún tipo de control, ni por panel ni semana de referencia.

Tabla 4.6
Distribución alternativa de 50 unidades primarias de muestreo

		PANEL				Total
		11	**12**	**19**	**20**	
SEMANA	1	1	1	1	1	4
	2	1	1	1	1	4
	3	1	1	1	1	4
	4	1	1	1	1	4
	5	1	1	1	1	4
	6	2	1	1	1	5
	7	1	2	1	1	5
	8	1	1	1	1	4
	9	1	1	1	1	4
	10	1	1	1	1	4
	11	1	1	1	1	4
	12	1	1	1	1	4
Total		**13**	**13**	**12**	**12**	**50**

La tabla 4.6 muestra una opción para el balanceado de la muestra de 50 unidades primarias de muestreo. Se puede apreciar un balanceado correcto tanto para paneles como semanas de referencia.

4.3 Estimadores y sus varianzas

En la aplicación de los estimadores se trabajará con la sub muestra de la ciudad de La Paz correspondiente al primer y segundo trimestre 2010.

El cálculo de los estimadores de cambio, estimador de promedio y estimador combinado se realiza bajo el supuesto de una muestra aleatoria simple monoetápica de viviendas.

La tabla 4.7 y la Tabla 4.8 muestran la distribución de viviendas por panel y semanas de referencia del primer y segundo trimestre respectivamente.

Tabla 4.7
Distribución de viviendas en la muestra proporcionada del 1er trimestre 2010

		PANEL				
		11	**12**	**19**	**20**	**Total**
SEMANA	1	0	8	0	26	34
	2	17	8	24	0	49
	3	18	8	22	6	54
	4	8	9	8	14	39
	5	21	16	8	0	45
	6	8	0	9	8	25
	7	7	0	7	0	14
	8	8	15	7	16	46
	9	8	8	0	9	25
	10	0	6	7	0	13
	11	9	0	0	0	9
	12	8	7	9	17	41
Total		**112**	**85**	**101**	**96**	**394**

Tabla 4.8
Distribución de viviendas en la muestra proporcionada 2do trimestre 2010

		PANEL				
		11	**15**	**20**	**21**	**Total**
SEMANA	1	8	0	26	0	34
	2	17	6	0	0	23
	3	35	6	7	7	55
	4	9	0	16	0	25
	5	23	13	16	8	61
	6	7	7	9	0	25
	7	6	8	9	0	23
	8	8	12	17	0	38
	9	8	8	8	0	24
	10	9	5	8	0	23
	11	17	17	0	0	34
	12	9	9	16	0	34
Total		**156**	**91**	**132**	**15**	**394**

Para calcular el traslape de la muestra se trabaja con el identificador único de vivienda, que en la base tiene el nombre de folio_viv. Para que una vivienda pertenezca al traslape debe figurar tanto en el primer trimestre como en el segundo trimestre 2010. Con el programa estadístico SPSS se emparejó la base usando la variable folio_viv, y se halló una sola base con información de las viviendas en el 1er trimestre y 2do trimestre[10].

La descripción de la base de datos traslapada en términos de viviendas se ve en la Tabla 4.9.

La base de 394 viviendas que se tiene en cada trimestre, es reducida a una sola base de datos con 192 viviendas, las cuales formaron parte de la muestra tanto en el primer trimestre como en el segundo trimestre.

[10] El programa utilizado para traslapar la base de datos está en el anexo 4.1

Tabla 4.9
Distribución de la sub muestra en el traslape 1er y 2do trimestre 2010

		PANEL		
		11	**20**	**Total**
SEMANA	1	0	25	25
	2	17	0	17
	3	16	6	22
	4	8	13	21
	5	18	0	18
	6	7	8	15
	7	5	0	5
	8	7	14	21
	9	8	8	16
	11	8	0	8
	12	8	16	24
Total		**102**	**90**	**192**

4.3.1 Estimador de cambio

En el cálculo del estimador de cambio se tomará la variable: número de personas por vivienda que pertenecen a la población económicamente activa.

Sea la variable:

$\hat{t}_i$: Número de personas por vivienda que pertenecen a la población económicamente activa (PEA[11]) en el trimestre i del año 2010

Se desea calcular el cambio que tuvo el promedio de la variable del número de personas que pertenecen a la PEA por vivienda del 1er trimestre al 2do trimestre 2010.

[11] Población económicamente activa es una subdivisión de la población en edad de trabajar. (Cartilla Metodológica de la Encuesta Trimestral de Empleo 2009)

Sea:

$\hat{\bar{t}}_i$: Promedio de personas por vivienda que pertenecen a la PEA en el trimestre i del año 2010

El estimador de cambio está definido por:

$$\hat{\theta} = \hat{\bar{t}}_2 - \hat{\bar{t}}_1$$

Cuya varianza es:

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

Por la ecuación (3.5) de la página 40 se tiene:

$$Var(\hat{\theta}) = Var(\hat{\bar{t}}_2) + Var(\hat{t}_1) - 2\beta_{1,2}\rho_{1,2}\sqrt{Var(\hat{t}_1)Var(\hat{t}_2)}$$

Asumiendo una población infinita y un muestreo aleatorio simple, el estimador de la varianza en cualquier ronda es:

$$Var(\hat{\bar{t}}_i) = \frac{\sigma_i^2}{n} \qquad i = 1,2$$

Donde n es el tamaño de muestra constante en cualquier trimestre

Se realiza un análisis descriptivo para el 1er trimestre y 2do trimestre de la variable “número de personas en la población económicamente activa por vivienda” y se obtiene los resultados que muestran la Tabla 4.9 y la Tabla 4.10

Tabla 4.10
Estadísticas descriptivas 1er Trimestre 2010

Variable $\hat{t}_i$		Mínimo	Máximo	Media		Desviación Std.	Varianza
	Tamaño de muestra n	Estadística	Estadística	Estadística	Error Estándar	Estadística	Estadística
PEA / VIV	394	,00	7,00	1,9594	,06506	1,29134	1,668

Tabla 4.11
Estadísticas descriptivas 2do Trimestre 2010

Variable $\hat{t}_i$		Mínimo	Máximo	Media		Desviación Std.	Varianza
	Tamaño de muestra n	Estadística	Estadística	Estadística	Error Estándar	Estadística	Estadística
PEA / VIV	394	,00	7,00	1,8883	,06137	1,21806	1,484

La matriz de correlación para la misma variable se calcula con la base traslapada (ver Anexo 4.1) y es:

Tabla 4.12
Matriz de Correlación

Correlación			
		PEA Trim 1	PEA Trim 2
PEA Trim 1	Correlación Pearson	1	,829**
	n	192	192
PEA Trim 2	Correlación Pearson	,829**	1
	n	192	192

Estimador de cambio para la muestra panel

Con los datos calculados, el estimador de cambio del segundo trimestre respecto del primer trimestre se calcula bajo la siguiente expresión:

$$\hat{\theta} = \hat{\bar{t}}_2 - \hat{\bar{t}}_1$$

$$\hat{\theta} = 1.8883 - 1.9594$$

$$\hat{\theta} = -0.0711$$

El signo negativo en el estimador de cambio indica que el promedio ha tenido una disminución con respecto al primer trimestre.

En el cálculo de la varianza influye la información del panel, específicamente en el coeficiente de correlación. Utilizando la expresión de la ecuación (3.5) se tiene lo siguiente:

$$Var_{con.rot}(\hat{\theta}) = Var(\hat{\bar{t}}_2) + Var(\hat{t}_1) - 2\beta_{1,2}\rho_{1,2}\sqrt{Var(\hat{t}_1)Var(\hat{t}_2)} \tag{4.1}$$

Con los datos obtenidos de la Tabla 4.9 se tiene:

$$\beta_{1,2} = {}^{192}\!/_{394} = 0.4873$$

$$\rho_{1,2} = 0.829$$

Reemplazando los valores en la ecuación (4.1):

$$Var_{con.rot}(\hat{\theta}) = 0.004232 + 0.003766 - 2 \times 0.4873 \times 0.829 \times \sqrt{0.04232 \times 0.003766}$$

$$Var_{con.rot}(\hat{\theta}) = 0.004772$$

Estimador de cambio para muestras independientes

Bajo el supuesto de muestras independientes el estimador de cambio se calcula de la misma forma que se calculó para el caso de muestras panel, esto es:

$$\hat{\theta} = \hat{\bar{t}}_2 - \hat{\bar{t}}_1$$
$$\hat{\theta} = 1.8883 - 1.9594$$
$$\hat{\theta} = -0.0711$$

En el cálculo de la varianza no se toma en cuenta la covarianza puesto que se está bajo el supuesto de muestras independientes.

$$Var_{Sin.rot}(\hat{\theta}) = Var(\hat{\bar{t}}_2) + Var(\hat{t}_1)$$

$$Var_{Sin.rot}(\hat{\theta}) = 0.004232 + 0.003766$$

$$Var_{Sin.rot}(\hat{\theta}) = 0.00799$$

La eficiencia del estimador obtenido mediante muestra panel con respecto a una muestra independiente es:

$$\frac{Var_{con.rot}(\hat{\theta})}{Var_{\sin.rot}(\hat{\theta})} = \frac{0.004772}{0.007990} = 0.5971$$

Es claro que la varianza del estimador de cambio calculada con muestras panel es menor que la varianza obtenida mediante muestras independientes. Esto se ve reflejado en el cociente de las varianzas cuyo resultado es 0.5971, que se interpreta como una reducción del 40% aproximadamente respecto a la varianza obtenida muestras independientes.

4.3.2 Estimador para el promedio

Si se desea obtener un estimador semestral de personas en la PEA por vivienda, este cálculo se puede realizar mediante un promedio simple de dos trimestres. Por el hecho que se está trabajando con muestras traslapadas, es necesario considerar una correlación de la variable entre los dos trimestres.

Sea el estimador semestral del promedio:

$$\hat{\theta}_p = \frac{\hat{\bar{t}}_1 + \hat{\bar{t}}_2}{2}$$

Según la ecuación (3.10), este estimador tiene varianza igual a:

$$Var(\hat{\theta}_p) = Var(\hat{t}_{i+1}) + Var(\hat{t}_i) + 2\beta_{i,i+1}\rho_{i,i+1}\sqrt{Var(\hat{t}_i)Var(\hat{t}_{i+1})}$$

Para muestras panel

El estimador semestral se puede calcular mediante el promedio de los dos trimestres.

$$\hat{\theta}_p = \frac{\hat{\bar{t}}_1 + \hat{\bar{t}}_2}{2}$$

$$\hat{\theta}_p = \frac{1.9594 + 1.8883}{2} = 1.9239$$

La varianza del estimador semestral es:

$$Var_{con.rot}(\hat{\theta}_p) = \frac{1}{4}\left[Var(\hat{t}_1) + Var(\hat{t}_2) + 2\beta_{1,2}\rho_{1,2}\sqrt{Var(\hat{t}_1)Var(\hat{t}_2)}\right]$$

$$Var_{con.rot}(\hat{\theta}_p) = \frac{1}{4}\left[0.004232 + 0.003766 + 2 \times 0.4873 \times 0.829\sqrt{0.004232 \times 0.003766}\right]$$

$$Var_{con.rot}(\hat{\theta}_p) = 0.002806$$

Para muestras independientes

El estimador semestral se calcula de igual forma que para el caso anterior

$$\hat{\theta}_p = \frac{\hat{\hat{t}}_1 + \hat{\hat{t}}_2}{2}$$

$$\hat{\theta}_p = \frac{1.9594 + 1.8883}{2} = 1.9239$$

La varianza del estimador semestral es:

$$Var_{con.rot}(\hat{\theta}_p) = \frac{1}{4}\left[Var(\hat{t}_1) + Var(\hat{t}_2)\right]$$

$$Var_{con.rot}(\hat{\theta}_p) = \frac{1}{4}\left[0.004232 + 0.003766\right]$$

$$Var_{con.rot}(\hat{\theta}_p) = 0.001999$$

La eficiencia del estimador obtenido mediante muestra panel con respecto a una muestra independiente es:

$$\frac{Var_{con.rot}(\hat{\theta}_p)}{Var_{\sin.rot}(\hat{\theta}_p)} = \frac{0.002806}{0.001999} = 1.4033$$

El hecho que la razón de las dos varianzas de los estimadores sea un número mayor a uno, indica que el estimador mediante muestras independientes es mejor que el estimador mediante muestras panel. En este caso se puede interpretar que con una muestra panel se obtiene una varianza incrementada en un 40% respecto a la varianza obtenida mediante muestras independientes.

4.3.3 Estimador Combinado

Ahora calcularemos el estimador combinado propuesto en el documento, el cual disminuye la varianza de una ronda dada, utilizando información de una ronda anterior.

Veamos los resultados que se tienen después de realizar un análisis descriptivo de las bases de datos.

Tabla 4.13

Estadisticas Descriptivas - Traslape 1er Trimestre							
	Tamaño de muestra n	Mínimo	Máximo	Promedio		Desviación Std.	Varianza
	Estadística	Estadística	Estadística	Estadística	Error Estándar	Estadística	Estadística
PEA / VIV	192	,00	6,00	1,9271	,09179	1,27189	1,618

La Tabla 4.13 muestra el análisis descriptivo de la variable: Número de personas en la PEA por vivienda del 1er trimestre. El análisis corresponde a las viviendas que se traslapan con el segundo trimestre. El total de viviendas traslapadas es 192 para ambos trimestres.

Tabla 4.14

Estadísticas Descriptivas - No Traslape 1er Trimestre							
	Tamaño de muestra n	Mínimo	Máximo	Promedio		Desviación Std.	Varianza
	Estadística	Estadística	Estadística	Estadística	Error Estándar	Estadística	Estadística
PEA / VIV	202	,00	7,00	1,9901	,09231	1,31198	1,721

La Tabla 4.14 muestra el análisis descriptivo de la variable: Número de personas en la PEA por vivienda del 1er trimestre, correspondiente a las viviendas que no pertenecen al traslape. El total de viviendas es 202 para ambos trimestres.

Tabla 4.15

Estadisticas Descriptivas - Traslape 2do Trimestre							
	Tamaño de muestra n	Mínimo	Máximo	Promedio		Desviación Std.	Varianza
	Estadística	Estadística	Estadística	Estadística	Error Estándar	Estadística	Estadística
PEA / VIV	192	,00	6,00	1,9688	,09251	1,28179	1,643

La Tabla 4.15 muestra el análisis descriptivo de la variable: Número de personas en la PEA por vivienda del 2do trimestre, correspondiente a las viviendas que pertenecen al traslape del 1er y 2do trimestre. Al igual que en el primer trimestre, el total de viviendas traslapadas es 192.

Tabla 4.16

Estadisticas Descriptivas - No Traslape 2do Trimestre							
	Tamaño de muestra n	Mínimo	Máximo	Promedio		Desviación Std.	Varianza
	Estadística	Estadística	Estadística	Estadística	Error Estándar	Estadística	Estadística
PEA / VIV	202	,00	7,00	1,8119	,08107	1,15222	1,328

Finalmente, la Tabla 4.16 muestra el análisis descriptivo de la variable: Número de personas en la PEA por vivienda del 2do trimestre. Este análisis corresponde a las viviendas que no pertenecen al traslape. El total de viviendas para el segundo trimestre es 202.

El estimador combinado propuesto para el segundo trimestre tiene la siguiente forma:

$$\hat{\hat{t}}_{C2} = \phi \cdot \hat{\hat{t}}_2^{n_2} + (1-\phi)\hat{\hat{t}}_2^{m_2}$$

Y varianza:

$$V(\hat{\hat{t}}_{C2}) = \phi^2 \cdot V(\hat{\hat{t}}_2^{n_2}) + (1-\phi)^2 V(\hat{\hat{t}}_2^{m_2})$$

Es claro que los estimadores $\hat{\hat{t}}_2^{n_2}$ y $\hat{\hat{t}}_2^{m_2}$ no tienen elementos en común, por lo tanto no existe covarianza entre ellos. Por eso la varianza del estimador combinado es simplemente la suma de las varianzas de los dos estimadores.

Donde:

- $\hat{\hat{t}}_2^{n_2}$: Estimador del promedio en el segundo trimestre de la parte no traslapada (Tabla 4.16)

$$\hat{\hat{t}}_2^{n_2} = \sum_{j=1}^{n_2} \frac{t_{j,2}}{n_2} = 1.8119$$

Y la varianza bajo el supuesto de un muestreo aleatorio simple:

$$V(\hat{\hat{t}}_2^{n_2}) = \frac{\sigma^2}{n_2} = 0.006572$$

- $\hat{\bar{t}}_2^{m_2}$: Estimador del promedio del segundo trimestre de la parte traslapada

$$\hat{\bar{t}}_2^{m_2} = \sum_{j=1}^{n} \frac{t_{j,1}}{n} + \left(\frac{1}{m_2}\sum_{j=1}^{m_2} t_{j,2} - \frac{1}{m_2}\sum_{j=1}^{m_2} t_{j,1}\right)$$

Donde:

$\sum_{j=1}^{n} \frac{t_{j,1}}{n} = 1.9594$: Estimador del promedio del 1er trimestre de la muestra total (ver Tabla 4.10)

$\frac{1}{m_2}\sum_{j=1}^{m_2} t_{j,2} = 1.9688$: Estimador del promedio del segundo trimestre solo con información del traslape (ver Tabla 4.15)

$\frac{1}{m_2}\sum_{j=1}^{m_2} t_{j,1} = 1.9271$: Estimador del promedio del 1er trimestre sólo con información del traslape (ver Tabla 4.13)

Reemplazando los datos en $\hat{\bar{t}}_2^{m}$:

$$\hat{\bar{t}}_2^{m_2} = \sum_{j=1}^{n} \frac{t_{j,1}}{n} + \left(\frac{1}{m_2}\sum_{j=1}^{m_2} t_{j,2} - \frac{1}{m_2}\sum_{j=1}^{m_2} t_{j,1}\right) = 1.9594+1.9688-1.9271$$

$$\hat{\bar{t}}_2^{m_2} = 2.0011$$

Y varianza:

$$V(\hat{\hat{t}}_2^{m_2}) = \frac{\sigma^2}{n\beta_2}\left[1+\alpha_2(1-2\rho_2)\right] = \frac{1.484}{394\times 0.487}\left[1+0.513\times(1-2\times 0.829)\right] = 0.005123$$

- ϕ: Ponderador calculado para minimizar la varianza del estimador combinado:

$$\phi = \frac{V(\hat{\hat{t}}_2^{m})}{V(\hat{\hat{t}}_2^{n_2}) + V(\hat{\hat{t}}_2^{m})} = \frac{0.005123}{0.006572+0.005123} = 0.4381$$

Ahora reemplazando estos datos en el estimador combinado del 2do trimestre de la ecuación (3.26) tenemos:

$$\hat{\hat{t}}_{C2} = \phi\cdot\hat{\hat{t}}_2^{n_2} + (1-\phi)\hat{\hat{t}}_2^{m_2} = 0.4381\times 1.8119 + (1-0.4381)\times 2.0011$$

$$\hat{\hat{t}}_{C2} = 1.9182$$

Y varianza:

$$V(\hat{\hat{t}}_{C2}) = \phi^2\cdot V(\hat{\hat{t}}_2^{n_2}) + (1-\phi)^2 V(\hat{\hat{t}}_2^{m_2}) = 0.4381^2\times 0.006572 + (1-0.4381)^2\times 0.005123$$

$$V(\hat{\hat{t}}_{C2}) = 0.002879$$

Comparando la varianza que se obtiene del estimador combinado con la varianza obtenida con el estimador común del segundo trimestre se tiene:

$$\frac{Var\ (\hat{\hat{t}}_{C2})}{Var\ (\hat{\hat{t}}_2)} = \frac{0.002879}{0.003766} = 0.7645$$

Nuevamente está claro que la varianza del estimador combinado es menor que la varianza del estimador común del segundo trimestre. Esta reducción de la varianza es de casi 24% con respecto a la varianza del estimador

4.3.3.1 Porcentaje Óptimo de Traslape

Supongamos que se desea re ajustar el modelo de rotación para obtener estimadores de cada ronda más precisos mediante el estimador combinado. Esto se realiza calculando el porcentaje óptimo mediante la ecuación (3.37) de la página que se halló en el capítulo anterior.

Con la información del coeficiente de correlación obtenida de la variable de interés de dos rondas consecutivas se calcula el porcentaje óptimo de traslape:

$$\beta_2 = \frac{2(1-\rho)-\sqrt{2-2\rho}}{1-2\rho} = \frac{2\times(1-0.829)-\sqrt{2-2\times 0.829}}{1-2\times 0.829}$$

$$\beta_2 = 0.369$$

Por lo tanto el porcentaje óptimo de traslape para minimizar aún más la varianza mediante el estimador combinado para un coeficiente de correlación igual a 0.829 es 37%. Esto quiere decir que si el principal objetivo es mejorar la precisión del estimador de una ronda, lo ideal es que el traslape de una ronda dada con la siguiente ronda debe reducirse el porcentaje de traslape de 48 % a 37%.

En la Tabla 3.3 de la página 63, vemos que el efecto de trabajar con el traslape óptimo produce una ganancia en precisión de 22% aproximadamente.

Conclusiones

1. Mediante la aplicación se pudo comprobar que las encuestas panel rotatorio con traslape parcial son útiles para calcular un estimador de cambio más precisamente que encuestas con muestras independientes. Con los datos proporcionados se probó que la varianza del estimador obtenido mediante muestras panel es más pequeña que la varianza del estimador con muestras independientes.

2. Como se ha visto en la aplicación, mediante las encuestas con diseño panel rotatorio con traslape parcial se puede mejorar el estimador de una ronda cualquiera utilizando información de una ronda anterior. Esta mejora se logra a través del estimador combinado.

3. En cambio, la varianza del estimador del promedio de dos o más rondas, calculado con muestras panel resultó ser mayor a la varianza obtenida mediante muestras independientes. Si esto representa un problema, y el objetivo es obtener estimadores de promedio de rondas consecutivas, entonces se debe optar por muestras independientes.

4. Cuando se trabaja con encuestas tipo panel rotatorio con traslape parcial, es importante aprovechar todas las bondades que brindan este tipo de encuestas. Por lo que vimos en el capítulo de aplicación, el INE podría mejorar los estimadores puntuales que generan, utilizando información de un periodo anterior. También podrían mejorar los estimadores de cambio aprovechando que se puede usar la covarianza entre rondas para disminuir la varianza. Tomar en cuenta que el estimador que se desarrolló en el capítulo de aplicación es el estimador del promedio, si se desea mejorar estimadores de razón, es necesario desarrollar toda la metodología del cálculo de estos estimadores, que seguramente serán más complejos.

5. Se presentan diferentes problemas al momento de diseñar la encuesta panel. Generalmente uno de los problemas más comunes es decidir por el modelo de rotación. En el documento se explicó las bondades de algunos modelos más comunes utilizados en las encuestas de varios países. De esta forma será más fácil decidir cuál de ellos usar para poder alcanzar los objetivos de planteados.

6. En el documento se planteó posibles soluciones a algunos problemas que generalmente se presentan al momento de realizar el diseño panel, estas son:

    - La forma más adecuada de iniciar algunos modelos de rotación para que poco a poco se vayan regularizando y finalmente el modelo de rotación sea totalmente regular.

    - Las reglas de selección a partir de la segunda ronda

    - El balanceado de la muestra en cuanto a paneles y tiempos de referencia

ANEXOS

ANEXO 2.1

Ficha metodológica Encuestas de Empleo en algunos países latinoamericanos[12]

Características	Bolivia	Chile	Colombia	Argentina
Año de implementación	**2011**	**2007**	**2006**	**2003**
Diseño de muestra	Panel trimestral Parcial 2-(2)-2	Panel 6 Urbano, Panel 9 Resto urbano, Panel 12 Rural , trimestrales	Encuesta Continua	Panel trimestral Parcial 2-(2)-2
	Estratificado bi etápico	Estratificado bi etápico		Estratificado bi etápico
Semanas / Trimestre	13 semanas	n.d.	13 semanas	12 semanas
	04/04/2005			04/04/2004
Periodicidad de entrega de resultados	Mensual, Trimestral, Anual	Trimestral, Anual, trimestre móvil	Mensual, Trimestral, semestral, Anual	Trimestral, semestre, Anual
Tamaño de muestra (Viviendas)	10.578 / Trim.	36.000 / año	271.620 /año	100.000/año
Tasa de no respuesta[13]	13%[2011]	10%[2006]	4% en principales ciudades [2000]	n.d.
Tasa de rechazo	5%	n.d.	n.d.	n.d.
Población objetivo	10 años y más	15 años y más	10 años y más (rural), 12 años y más años (urbano)	10 años y más
Área geográfica	Ciudades capitales + El Alto	Urbano - Rural	Urbano - Rural	Urbano - Rural
Participación del informante	Voluntario	Voluntario	Voluntario	Voluntario

[12] Extraído de documentos metodológicos disponibles en Portales de Institutos de Estadística respectivos
[13]Extraído del portal de la Organización Internacional del Trabajo (http://www.ilo.org/dyn/lfsurvey/)

ANEXO 2.2

Ficha metodológica Encuestas de Empleo en algunos países latinoamericanos[14]

Características	Perú	Ecuador	Brasil	Costa Rica
Año de implementación	**2001**	**2007**	**2001**	**2009**
Diseño de muestra	Panel trimestral Parcial 2-(2)-2	Panel trimestral Parcial 2-(2)-2	Panel mensual parcial 4	Panel trimestral 4
	Estratificado bi etápico	Estratificado tri etápico	Estratificado bi etápico	Estratificado bi etápico
Semanas / Trimestre	12 semanas	n.d.	365 días	n.d.
	04/04/2004			
Periodicidad de entrega de resultados	Trimestral, Anual, trimestre móvil	Trimestral, Anual	Mensual	Trimestral, Anual
Tamaño de muestra (Viviendas)	3.000 / Trim.	5.988 / Trim.	37.212/ Mes	9.024/ Trim.
Tasa de no respuesta[15]	12%[2000]	n.d.	2.5%[2007]	n.d.
Tasa de rechazo	n.d.	n.d.	n.d.	n.d.
Población objetivo	14 años y más	10 años y más	10 años y más	15 años y más
Área geográfica	Lima metropolitana y Callao	Urbano - Rural	Urbana 6 regiones metropolitanas	Urbano - Rural
Participación del informante	Voluntario	Voluntario	Voluntario	Voluntario

[14] Extraído de documentos metodológicos disponibles en Portales de Institutos de Estadística respectivos
[15] Extraído del portal de la Organización Internacional del Trabajo (http://www.ilo.org/dyn/lfsurvey/)

ANEXO 2.3

Ficha metodológica Encuestas de Empleo en algunos países europeos[16]

Características	Bélgica	Alemania	Finlandia	Sud África
Año de implementación	**1999**	**2005**	**2000**	**n.d.**
Diseño de muestra	Encuesta Continua	Panel anual 4	Panel trimestral 3-(1)-2	Panel trimestral 5
Semanas/Trimestre	52 semanas/ año	52 semanas/ año	13 semanas 04/04/2005	n.d.
Periodicidad de entrega de resultados	Trimestral, Anual	Trimestral, anual	Mensual, trimestral, anual	Mensual, trimestral, anual
Tamaño de muestra (Viviendas)	14.925/Trim.	-	-	30.000/Trim.
Tasa de no respuesta	23,70%	2%	20,20%	15%
Tasa de rechazo	3%	n.d.	10%	2%
Población objetivo	15- 75 años	15 años y más	15- 74 años	15-65 años
Área geográfica[17]	Todo el país	Todo el país	Todo el país	Todo el país
Participación del informante	Obligado	Obligado	Voluntario	n.d

[16] Extraído del documento de (http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/publication?p_product_code=KS-RA-09-002)

[17] Extraído del portal de la Organización internacional del Trabajo (http://www.ilo.org/dyn/lfsurvey/)

ANEXOS 2.4

Ficha metodológica Encuestas de Empleo en Sud África y algunos países europeos[18]

Características	Suecia	Reino Unido	España	Francia
Año de implementación	**1993**	**1992**	**1999**	**2003**
Diseño de muestra	Panel trimestral 8	Panel trimestral 5	Panel trimestral 6	Panel trimestral 6
	Estratificado simple	Estratificado mono etápico	Estratificado bi etápico	Estratificado bi etápico
Semanas/Trimestre	13 semanas	13 semanas	13 semanas	13 semanas
	04/04/2005	04/04/2005		
Periodicidad de entrega de resultados	Mensual, trimestral, anual	Mensual, trimestral, anual	Trimestral, Anual	Trimestral, Anual
Tamaño de muestra (Viviendas)	-	55.000/ Trim.	65.000/Trim.	56.000 /Trim.
Tasa de no respuesta	18,40%	30,4	18,80%	17%
Tasa de rechazo	9%	15%	6%	4%
Población objetivo	15- 74 años	16 años y más	15 años y más	15 años y más
Área geográfica[19]	Todo el país	Todo el país	Todo el país	Francia Metro-politana[2000]
Participación del informante	Voluntario	Voluntario	Obligado	Obligado

[18] Extraído del documento de (http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/publication?p_product_code=KS-RA-09-002)
[19] Extraído del portal de la Organización internacional del Trabajo (http://www.ilo.org/dyn/lfsurvey/)

ANEXO 3.1

Teorema 3.1

El valor esperado de cualquier variable aleatoria X se puede calcular:

$$E(X) = E(E(X/u_i) = E_1 E_2(X)$$

Con $u_i (i = 1,2,..,n)$ un conjunto de eventos mutuamente excluyentes

Demostración[20]

$$E(X) = \sum_i x_i p(X = x_i)$$

$$E(X) = \sum_i x_i \sum_j p(X = x_i, U = u_j)$$

$$E(X) = \sum_i x_i \sum_j p(X = x_i / u_j) p(u_j)$$

$$E(X) = \sum_j p(u_j) \sum_i x_i \cdot p(X = x_i / u_j)$$

$$E(X) = \sum_j p(u_j) E(X/u_j) = \sum_j p(u_j) E_2(X)$$

$$E(X) = E_1 E_2(X)$$

$E_2(X)$ es el valor esperado de X dado u_j, y *E1* Representa el procedimiento siguiente de tomar el valor esperado.

[20] Angulo A. (1991) Anexos

ANEXO 3.2

Teorema 3.2

$$\operatorname{cov}(X,Y)=E_1[\operatorname{cov}_2(X,Y)]+\operatorname{cov}_1[E_2(X),E_2(Y)]$$

Demostración[21]

$$\begin{aligned}
&\operatorname{cov}(X,Y)=E(X,Y)-E(X)E(Y)\\
&\operatorname{cov}(X,Y)=E_1E_2(X,Y)-E_1E_2(X)\cdot E_1E_2(Y)\\
&\operatorname{cov}(X,Y)=E_1[E_2(X,Y)-E_2(X)E_1(Y)]+E_1[E_2(X)E_2(Y)]-E_1E_2(X)\cdot E_1E_2(Y)\\
&\operatorname{cov}(X,Y)=E_1\operatorname{cov}_2(X,Y)+\operatorname{cov}_1[E_2(X),E_2(Y)]
\end{aligned}$$

Colorario 3.2

$$\begin{aligned}
&\operatorname{cov}(X,X)=E_1\operatorname{cov}_2(X,X)+\operatorname{cov}_1[E_2(X),E_2(X)]\\
&\operatorname{cov}(X,X)=E_1V_2(X)+V_1E_2(X)
\end{aligned}$$

Conocido como el teorema de Hansen, Hurvitz y Madow (1953)

[21] Angulo A. (1991) anexos

ANEXOS 4.1

Para calcular la matriz de correlación primeramente hay que tener una base traslapada para luego calcular dicha matriz. La variable con la que se puede emparejar los datos es “fol_miem”. El programa utilizados es:

```
GET
  FILE='C:\Users\Administrador\Desktop\tesis mercy\trim 1 2010.sav'.
DATASET ACTIVATE DataSet1.
MATCH FILES /FILE=*
  /TABLE='C:\Users\Administrador\Desktop\tesis mercy\trim 2 2010.sav'
  /RENAME (sel upm = d0 d1)
  /BY fol_miem
  /DROP= d0 d1.
EXECUTE.
SORT CASES BY folio_2 (A).
SORT CASES BY folio (A).
STRING  folio_viv (A9).
COMPUTE folio_viv=CHAR.SUBSTR(folio,1,9).
EXECUTE.
SORT CASES BY folio_viv (A).
* Identify Duplicate Cases.
SORT CASES BY folio_viv(A).
MATCH FILES
  /FILE=*
  /BY folio_viv
  /FIRST=PrimaryFirst
  /LAST=PrimaryLast.
DO IF (PrimaryFirst).
COMPUTE  MatchSequence=1-PrimaryLast.
ELSE.
COMPUTE  MatchSequence=MatchSequence+1.
END IF.
LEAVE  MatchSequence.
FORMAT  MatchSequence (f7).
COMPUTE  InDupGrp=MatchSequence>0.
SORT CASES InDupGrp(D).
MATCH FILES
  /FILE=*
  /DROP=PrimaryFirst InDupGrp MatchSequence.
VARIABLE LABELS  PrimaryLast 'Indicator of each last matching case as Primary'.
VALUE LABELS  PrimaryLast 0 'Duplicate Case' 1 'Primary Case'.
VARIABLE LEVEL  PrimaryLast (ORDINAL).
FREQUENCIES VARIABLES=PrimaryLast.
EXECUTE.
CORRELATIONS
  /VARIABLES= pea pea_2
  /PRINT=TWOTAIL NOSIG
  /MISSING=PAIRWISE
```

Referencias

Kasprzyk, D. , Duncan, G. , Kalton, G. & Singh, M. P. (1989). *Panel Surveys*.Canadá: Editorial Willey

Lynn P. (2005). *Longitudinal survey methodology*, recuperado en octubre 5, 2011 de http://www.eustat.es/productosServicios/datos/Sem45_c.pf

Angulo, A. (1991). *Muestreo en ocasiones sucesivas*. Tesis no publicada de estadística, Universidad Mayor de San Andrés, La Paz - Bolivia

Cochran, W.G. (1977). *Técnicas de Muestreo*, 3ra edición. New York

Department of Economic and Social Affairs Statistics Division Studies in Methods (2007). *Household Sample Surveys in Developing and Transition Countries*. Naciones Unidas. Autor

Bautista L. (1998). *Diseños de Muestreo Estadístico.* Departamento de Matemática y Estadística.

Técnicas de estudio. *Metodología de la investigación*. Recuperado en Septiembre 21, 2011 de http://www.tecnicas-de-estudio.org/investigacion/investigacion40.htm

Medina F. (Cepal), Encuestas panel . Recuperado en octubre 20, 2011 de http://www.eclac.cl/deype/mecovi/docs/TALLER7/12.pdf

García, D. , Nieto A. & Trapero, B. (s.a) *Estudio de la eficacia de un muestreosucesivo con reemplazamiento parcial, al utilizar un estimador de la media en lugar del estimador de regresión.* Estadísticos Facultativos

Eckler, A. (1955), The Annals of Mathematical Statistics, *Rotation Sampling, pp.* 664 – 665.

García, A. & Artés, E. (2002), *Aportaciones al muestreo sucesivo*, RevistaColombiana de Estadística, Volumen 25 Nº 2, pp. 145 - 148