

Diseño muestral y ponderación encuesta de Innovación 2016-2018

1. Condición de elegibilidad

Son elegibles para participar de la Encuesta de Innovación aquellas empresas cuyo personal ocupado promedio es mayor o igual a cinco y/o sus ventas son mayores a 120 millones de pesos; y su actividad económica principal según la CIIU Rev. 4 se encuentra comprendida en las siguientes secciones:

- C. Industrias Manufactureras
- D. Suministro de electricidad, gas, vapor y aire acondicionado
- E. Suministro de agua; alcantarillado, gestión de desechos y actividades de saneamiento (excluyendo las Divisiones 37 y 39)
- H. Transporte y almacenamiento
- I. Alojamiento y servicio de comida
- J. Información y comunicación
- K. Actividades financieras y de seguros (únicamente las Divisiones 64-66)
- M. Actividades profesionales, científicas y técnicas
- N. Actividades administrativas y servicios de apoyo, y Q. Servicios sociales y relacionados con la Salud Humana (excluyendo las Divisiones 87 y 88).

2. Diseño muestral

La muestra para el periodo 2016-2018 se encuentra compuesta por todas aquellas empresas elegibles respondientes en la ola anterior (2013 - 2015) y que continúan cumpliendo los criterios de elegibilidad en 2016 -2018, más un subconjunto de empresas nuevas que son seleccionadas en base al directorio de empresas del año 2018. Este subconjunto de empresas nuevas se le denomina de aquí en más, muestra de refrescamiento (*refresh*).

El *refresh* tiene como objetivo poder brindar utilizando la muestra completa (panel más *refresh*) estimaciones confiables para el periodo 2016-2018 (estimaciones transversales) mientras reduce el impacto producto de la atrición (no respuesta acumulativa) que sufren todos los paneles, así como reducir el efecto cohorte que sufren todos los paneles puros (sin *refresh*).

La muestra de *refresh* es seleccionada utilizando el mismo diseño muestral de las ediciones anteriores, es decir, un diseño directo, aleatorio y estratificado.

Los estratos son contruidos utilizando distintos niveles de información. En un primer nivel, las empresas son agrupadas en base a su actividad económica principal a nivel de división de actividad según la CIIU REV 4. Posteriormente, dentro de cada división, las empresas son

clasificadas en cuatro estratos de tamaño en base al Personal Ocupado (PO) promedio en el año. Los cuatro estratos de tamaño quedan definidos de la siguiente forma:

1. Empresas con PO entre 5 y 9.
2. Empresas con PO entre 10 y 19
3. Empresas con PO entre 20 y 49
4. Empresas con PO entre 50 y 99
5. Empresas con PO mayor o igual a 100 o sus ventas declaradas ante la DGI son mayores de 120 millones, a excepción de aquellas empresas con actividad económica principal es comercio, en donde, el límite inferior de las ventas para ser incluidas en este estrato se situó en 240 millones de pesos.

Los tamaños de muestra del refresh por estrato son asignados teniendo en cuenta dos factores:

- I. el tamaño de muestra efectivo en el panel
- II. el crecimiento del tamaño del estrato entre un periodo y otro.

3. Tamaño de muestra y tasa de efectividad.

El tamaño de muestra teórico para el período 2016-2018 se situó en 3070 empresas, de las cuales aproximadamente un 80 % corresponden a empresas del panel, es decir, empresas elegibles que respondieron en la edición anterior de la encuesta.

El tamaño de muestra efectivo se situó en 2978 lo que se traduce en una tasa de efectividad o logro de 97 % aproximadamente. Cabe aclarar que la tasa de efectividad no tiene porque coincidir con la tasa de respuesta, ya que la misma incluye empresas que una vez contactadas las mismas no cumplían los criterios de ser elegibles para participar del estudio. A su vez, de las empresas elegibles respondientes en el periodo 2016-2018, 2061 corresponden a empresas del panel.

4. Ponderación

Teniendo en cuenta el tipo de encuesta implementando, se definieron dos sistemas de ponderadores: transversales y longitudinales.

Los ponderadores transversales son utilizados para realizar estimaciones de los niveles de los distintos indicadores que aborda la encuesta para el periodo 2016-2018. Mientras que los ponderadores longitudinales tiene como objetivo poder medir cambios brutos (e.g. construcción de matrices de transición) para distintos indicadores del periodo de la edición anterior (2013-2015) y la esta última edición (2016-2018).

4.1. Ponderadores transversales

Para la construcción de los ponderadores transversales se tiene en cuenta toda la muestra, es decir, el panel más el refresh. Estos ponderadores son utilizados para realizar estimaciones ya

sea a nivel global o para distintos dominios de estimación de una amplia gama de indicadores para el periodo 2016-2018. El ponderador transversal para la empresa i que cumple los criterios de elegibilidad para participar del estudio y que respondió a la encuesta viene dado por:

$$w_i = d_{i0} \times nr_i \times g_i$$

donde d_{i0} es el ponderador original o basado en el diseño para la empresa i , nr_i es el ajuste por no respuesta y g_i es el ajuste proveniente de la calibración. A continuación se presentan con más detalle cada uno de los componentes definidos anteriormente

Ponderador original o basados en el diseño

El ponderador original viene determinado por la estrategia de selección de la muestra. Más precisamente viene determinado por el inverso de la probabilidad de inclusión o selección en la muestra. Teniendo en cuenta que dentro de cada estrato, las empresas son seleccionadas con igual probabilidad de selección, el ponderador original de la empresa i viene dado por:

$$d_{i0} = \frac{N_h}{n_{hp} + n_{hf}}$$

donde N_h es la cantidad de empresas en el marco muestral del año 2018 pertenecientes al estrato h , n_{hp} es la cantidad de empresas pertenecientes al panel para el estrato h y n_{hf} es el tamaño de muestra de refuerzo seleccionado en el estrato h para el periodo 2016-2018.

Ajuste por no respuesta.

La no respuesta se debe principalmente a deficiencias en el marco muestral en lo concerniente a las variables de identificación y ubicación de las empresas. La existencia de no respuesta implica que los ponderadores originales deban ser ajustados para corregir la subestimación provocada por la misma. El ajuste por no respuesta fue realizado a nivel de estrato y queda definido por el inverso de la tasa de respuesta en dicho estrato.

Ajuste por calibración

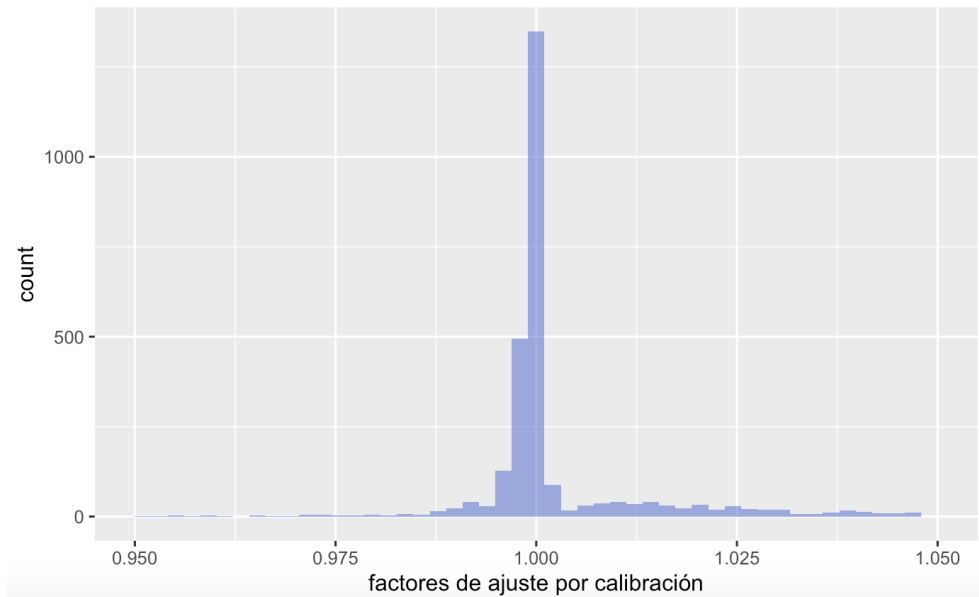
En este último paso los ponderadores son calibrados a información conocida del marco muestral de forma que las estimaciones provenientes de la encuesta coinciden de forma exacta con información del marco muestral.

Los principales beneficios de la calibración son:

1. Reducción de los errores estándar (SE) de las estimaciones de los distintos indicadores. Esto ocurre si las variables auxiliares de alguna forma explican la variabilidad en las variables de interés de la encuesta, es decir, se encuentran correlacionadas.
2. Posible reducción del sesgo por problemas de cobertura
3. Reducción del Sesgo ocasionado por la no respuesta. Esto ocurre si las variables auxiliares utilizadas para la calibración explican de alguna forma la propensión de responder de la empresa.

La información utilizada para la calibración es el total de empleados y cantidad de empresas a nivel de sección de actividad. Lo anterior, implica que las estimaciones de la cantidad de empresas y el total de empleados para cada una de las secciones de actividad coincidan con el marco muestral. Una vez definido el set de variables auxiliares para la calibración y definido los totales poblacionales, se utilizó la distancia *raking* para definir la modificación entre los ponderadores ajustados por no respuesta y los ponderadores finales. En la figura 1 se presenta la distribución de los factores de ajuste por calibración g_i

Figura 1: Distribución de los factores de ajuste por calibración de los pesos transversales



4.2 Ponderadores longitudinales

Para la construcción de los ponderadores o pesos longitudinales se tiene únicamente en cuenta la parte de la muestra del panel, es decir, aquellas empresas que participaron en las últimas dos olas del estudio. En las encuestas de panel, los ponderadores deben ser ajustados para tratar el problema de atrición, esto es, cuando una empresa que fue seleccionada en olas anteriores no responde en la ola actual.

El ajuste utilizado para el tratamiento de la no respuesta fue basado en un modelo de grupos de respuesta homogéneos. El mismo consiste en clasificar a las empresas elegibles de la muestra panel en C grupos para los cuales se supone que la probabilidad de responder son similares (e.i. presentan las mismas características). Los grupos fueron contruidos en base a la clasificación cruzada de la sección de actividad económica principal y el estrato de tamaño al cual pertenece la empresa. El ponderador de la empresa i que respondió en el periodo 2016-2018 perteneciente a una clase cualquiera c ajustado por no respuesta queda definido como: C grupos para los cuales se supone que la probabilidad de responder son similares (e.i. presentan las mismas características). Los grupos fueron contruidos en base a la clasificación cruzada de la sección de actividad económica principal y el estrato de tamaño al cual pertenece la empresa. El ponderador de la empresa i que respondió en el periodo 2016-2018 perteneciente a una clase cualquiera c ajustado por no respuesta queda definido como:

$$w_{i,t}^{\text{nr}} = \left(\sum_{i \in s_{c,E}} w_{i,t-1} \right) \times \left(\sum_{i \in s_{c,ER}} w_{i,t-1} \right)^{-1} \times w_{i,t-1}$$

donde $w_{i,t-1}$ es el ponderador transversal de la ola anterior (2013-2015) el cual fue ajustado por no respuesta y calibrado a conteos poblacionales provenientes del marco muestral, $s_{c,E}$ es el subconjunto de la muestra que respondieron en la periodo 2013-2015 pertenecientes al grupo o post estrato de no respuesta c y $s_{c,ER}$ es el subconjunto de empresas que contestaron en el periodo actual, es decir, 2016-2018.

Figura 2: Pesos transversales de la ola anterior (2013-2015) vs los pesos longitudinales de la ola (2016-2018)

