



Resultados del diagnóstico y prueba piloto de los indicadores de calidad aprobados de la fase I

Comité de Aseguramiento de la Calidad



Cuarta Sesión
Diciembre 12, 2017

Indicadores de precisión estadística



Indicadores de precisión y confiabilidad

- Diagnóstico:
 - Coeficiente de Variación (CV)
 - Cobertura de la variable de diseño para proyectos con muestreo no probabilístico
 - Tasa de no respuesta (TNR) a nivel de unidad de observación
- Piloto:
 - Tasa de no respuesta (TNR) para cada variable principal



CV: Diagnóstico

- Objetivo:

Analizar el conjunto de valores de los CVs asociados a las variables y desagregaciones para las cuales se realizó el diseño de cada uno de los proyectos, es decir, los CVs asociados a aquellas estimaciones en donde su precisión debería satisfacer **por diseño** un nivel pre-definido como aceptable.

- Actividades realizadas:

1. Especificación del conjunto de **variables principales** y **dominios de estudio** de cada proyecto.
2. Integración y análisis los coeficientes de variación calculados para cada proyecto.
3. Revisión de la experiencia internacional sobre los valores de referencia para su interpretación.

- Universo de estudio: 33 proyectos con muestreo probabilístico publicados a partir de 2016 (incluyendo aquellos levantados en 2015).

CV: Resultados

		Área administrativa					
		DGEE		DGES		DGEESPJ	
Número de proyectos		10		15		8	
Núm total de valores del CV*		394		3,635		990	
Mínimo		0.6%		0.0%**		0.3%	
Máximo		60.9%		99.5%***		48.4%	
		Acumulado		Acumulado		Acumulado	
Porcentaje de valores del CV	<15%	80.0%	80.0%	84.5%	84.5%	93.7%	93.7%
	[15%, 20%)	9.8%	89.8%	8.0%	92.5%	4.1%	97.8%
	[20%, 25%)	3.3%	93.1%	1.8%	94.3%	1.5%	99.3%
	[25%,30%)	3.1%	96.2%	0.7%	95.0%	0.3%	99.6%
	>= 30%	3.8%	100.0%	5.0%	100.0%	0.4%	100.0%

* El número total de valores del CV se obtiene al considerar el total de proyectos y sus respectivos coeficientes de variación calculados para todas las variables principales y dominios de estudio. Por ejemplo, en los proyectos a cargo de la DGEE el número de variables principales definidas por proyecto es igual 1 o 2 y el número de dominios de estudio va desde 1 y hasta 40.

** Valores de 0% fueron calculados para el Módulo de Ciber Acoso (MOCIBA) a nivel entidad federativa.

*** El CV de 99.5% corresponden a la estimación asociada a la variable “Otros Ingresos Corrientes” de la ENIGH 2016 en los dominios de estudio obtenidos con el cruce de entidad federativa y tipo de localidad.

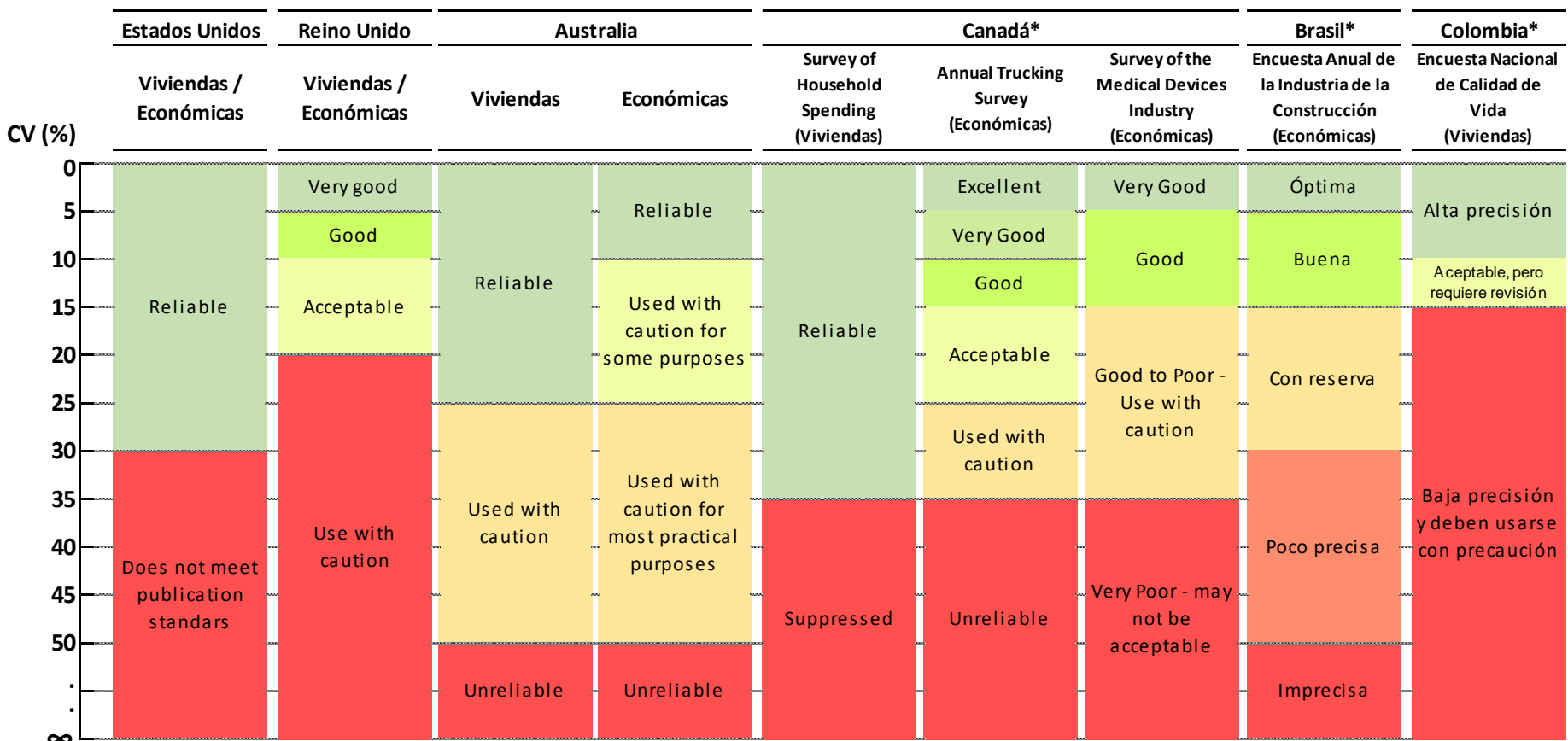
CV: Valores de referencia INEGI

Interpretación	Semaforización	Unidad de observación			
		Viviendas/Hogares DGES / DGE GSPyJ	Unidades económicas DGEE DGE GSPyJ		Otras (ENPOL) DGE GSPyJ
Buena		[0%, 15%)	[0%, 20%)	[0%, 15%)	[0%, 15%)
Aceptable		[15%, 25%)	[20%, 30%)	[15%, 30%)	[15%, 25%)
Con reserva		>=25%	>=30%	>=30%	>=25%

- Semaforización: en algunos proyectos se colorean las estimaciones en los tabulados con base en los valores de referencia definidos para el CV.
- Se manejan distintos valores dependiendo de la unidad de observación.
- Incluso considerando una misma unidad de observación, hay diferencias en los valores de referencia entre las áreas administrativas.

CV: Valores de referencia internacionales

- Existe heterogeneidad en los valores de referencia utilizados en otros países.
 - EEUU y Reino Unido: aplican los mismos valores para todo tipo de encuestas; EEUU tiene sólo un corte.
 - Australia: los valores de referencia sólo dependen de la unidad de observación.
 - Canadá, Brasil y Colombia: valores de referencia diferentes entre proyectos y por unidad de observación.
- El valor usado para alertar al usuario cuando la estimación es poco precisa varía entre 15 y 35%



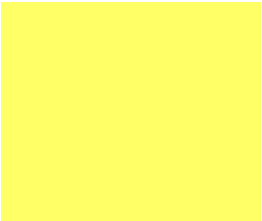

* Para estos países no se usa un conjunto de valores de referencia único, aquí sólo se presentan algunos ejemplos de encuestas particulares.

CV: Conclusiones

- Es factible calcular de manera estandarizada el CV en todos los **proyectos con muestreo probabilístico**, en particular para las variables principales y dominios de estudio.
- El uso del CV debe considerarse desde dos perspectivas:
 1. **Para informar sobre la calidad del proyecto.** Sólo se debería considerar el CV asociado a **variables principales y dominios de estudio** definidos en el diseño (podría implicar una revisión del diseño).
 2. **Para facilitar el uso correcto de la información a los usuarios.** Se podría usar el CV en todas las estimaciones junto con valores de referencia que orienten su interpretación, sobre todo cuando se realicen estimaciones **para otras variables y desagregaciones no consideradas en el diseño** pero que son de interés para el usuario.

CV: Conclusiones

- Pese a que uno de los objetivos era definir un conjunto único de valores de referencia para todos los proyectos, no se llegó a un acuerdo, por lo que se presentan dos propuestas:

<u>Precisión</u>	<u>Semaforización</u>	<u>Propuesta 1</u> DGES / DGE GSPyJ	<u>Propuesta 2</u> DGEE	<u>Texto de guía al usuario</u>
Alta		[0%, 15%)	[0%, 20%)	La precisión es alta y se puede usar la información con confianza.
Media		[15% , 30%)	[20% , 30%)	La precisión es aceptable para la mayoría de los propósitos, sin embargo, se sugiere al usuario que realice una revisión de otros indicadores para validar su uso, por ejemplo, el Intervalo de Confianza y el Error Estándar.
Baja		>=30%	>=30%	Este nivel de precisión de la estimación requiere un uso cauteloso de la información en el cual se analicen las causas de la alta variabilidad.

Cobertura de la variable de diseño: Diagnóstico

- Objetivo:

Analizar la cobertura asociada a las variables y desagregaciones para las cuales se realizó el diseño de cada uno de los proyectos con muestreo no probabilístico.

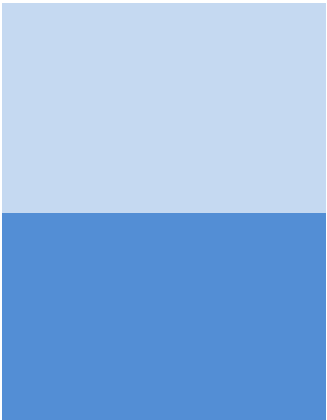
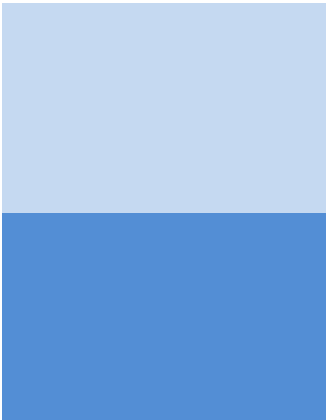
- Actividades realizadas:

1. Especificación del conjunto de **variables de diseño** y **dominios de estudio** de cada proyecto.
2. Integración y análisis de la tasa de cobertura.
3. Revisión de los valores de referencia utilizados actualmente en el INEGI para la interpretación de este indicador.

- Universo de estudio:

- 7 proyectos con muestreo no probabilístico publicados a partir de 2016 (incluyendo aquellos levantados en 2015). Todos los proyectos de este tipo están a cargo de la DGEE.

Cobertura de la variable de diseño: Resultados y valores de referencia INEGI

Nivel de Cobertura	Valores de Referencia INEGI			Resultados del diagnóstico. % de tasas de cobertura*
	Semaforización	Rango	Interpretación	
Alta		80% - 100%	Si el nivel de cobertura de la variable de diseño es mayor o igual al 80%, es considerada Alta.	63.30%
Media		60% - 80%	Si el nivel de cobertura de la variable de diseño es mayor o igual al 60% y menor al 80%, es considerada Media.	28.10%
Baja		< 60%	Si el nivel de cobertura de la variable de diseño es menor al 60%, es considerada Baja.	8.60%

El cálculo de este indicador sólo aplica a proyectos a cargo de la DGEE y en todos los casos el marco de referencia usado para el cálculo es “CE2009 ACT”.

* El porcentaje de tasas de cobertura en cada uno de los rangos se calcula sobre el número total de valores del indicador, el cual se obtiene al considerar el total de proyectos y sus respectivas tasas de cobertura calculadas para todas las variables de diseño y dominios de estudio. En los 7 proyectos la única variable de diseño fue “INGRESOS” y lo que hace la diferencia son los dominios de estudio.

Cobertura de la variable de diseño: Conclusiones

- Es factible calcular de manera estandarizada la cobertura de la variable de diseño para todos los dominios de estudio en todos los **proyectos con muestreo no probabilístico**.
- Los valores de referencia usados actualmente por el INEGI parecen adecuados para ayudar a los usuarios a interpretar el resultado de este indicador.



TNR a nivel unidad de observación: Diagnóstico

- Objetivo:
 - Verificar que las especificaciones (por ejemplo, la fórmula) incluidas en la ficha técnica del indicador aprobado por el CAC fueran suficientes para estandarizar su cálculo en cada uno de los diferentes tipos de proyecto.
- Actividades realizadas:
 1. Definición de los proyectos e identificación de las **unidades de observación, variables principales y agrupaciones** para realizar el cálculo del indicador en cada proyecto.
 2. Integración y análisis de las TNR calculadas para cada proyecto.
- Universo de estudio:
 - Cada área analizó al menos un proyecto con muestreo.
 - Adicionalmente, la DGES y la DGEE incluyeron un proyecto censal y un registro administrativo.



TNR a nivel unidad de observación: Resultados

- Se encontraron variaciones entre las UA en la conceptualización, uso y etapa del proceso en la que se calcula la TNR.
 - En los **proyectos con muestreo**, existe heterogeneidad en cuanto a la eliminación de toda la información de una unidad de observación debido a una respuesta parcial, esto se decide en algunos casos en campo y en otros después de aplicar criterios de validación en el sistema.
 - Es necesario clarificar la definición de No Respuesta y distinguirla del concepto de subcobertura, particularmente para el caso de **registros administrativos**.
 - Se requiere verificar el cálculo estandarizado de la no respuesta en los proyectos **censales**, ya que podría suceder lo mismo que en las encuestas.



TNR a nivel unidad de observación: Adecuaciones a la ficha técnica

- Con base en los resultados del diagnóstico en los **proyectos con muestreo probabilístico y muestreo no probabilístico**, se proponen las siguientes modificaciones a la ficha aprobada por el CAC el 30 de octubre de 2017.
 - Modificar el nombre del indicador a **“Tasa de no respuesta antes de imputación a nivel unidad de observación”** para explicitar el momento al cuál se refiere y en el que se debe calcular el indicador.
 - Se precisan el texto y fórmula, utilizando las fases del Modelo del Proceso Estadístico y Geográfico (MPEG), para precisar el momento en el que se tomarían los insumos del cálculo del indicador.



TNR a nivel unidad de observación: Conclusiones

- Es factible calcular de manera estandarizada la Tasa de no respuesta antes de imputación a nivel unidad de observación en todos los **proyectos con muestreo probabilístico y no probabilístico**.
 - El cálculo se realizará siguiendo la nueva ficha técnica.
- Se reprograma la presentación de resultados del diagnóstico para **censos y registros administrativos** para la primera sesión del 2018.
- Posteriormente, el grupo de trabajo analizará la conveniencia de agregar un conjunto de indicadores complementarios propuestos por la DGEE como parte de la fase II:
 - TNR mínima
 - TNR real (después de imputaciones)
 - Tasa de imputación



TNR para cada variable principal: Prueba piloto

- Motivación del indicador

Analizar el sesgo potencial en las estimaciones derivado de la eliminación de **información parcial** que se capta en algunas encuestas del INEGI, pero que no se incluye en la Base de Datos Pública de la encuesta, ni se utiliza para generar estimaciones. En algunos casos se elimina información parcial cuando una sola pregunta no se contesta de forma admisible.

- Objetivos:

1. Revisar la factibilidad del cálculo de la TNR para cada variable principal conforme a la fórmula de la ficha técnica aprobada por el CAC el 30 de octubre de 2017.
2. Revisar la aplicabilidad de las especificaciones de la fórmula para los cuatro tipos de proyectos (encuestas con muestreo probabilístico, muestreo no probabilístico, censos y registros administrativos)



TNR para cada variable principal: Prueba piloto

- Actividades realizadas:
 1. Definición de los proyectos e identificación de las **unidades de observación, variables principales y agrupaciones** para realizar el cálculo del indicador en cada proyecto.
 2. Integración y análisis de las TNR para cada una de las variables principales calculadas en cada proyecto.
 3. Elaboración de una nota explicativa en caso de no poder calcular el indicador.
- Proyectos analizados:
 - **DGEE**: CE2014 (Censo) y EMEC (Muestreo probabilístico y no probabilístico).
 - **DGES**: CPV2010 (censo), Defunciones Registradas (Registro administrativo) y EIC2015 (muestreo probabilístico). Aunque no se obtuvo el cálculo, también se analizó la factibilidad en otros proyectos con muestreo probabilístico, incluyendo ENOE, MTI, ENIGH, MCS, ENCO, BIARE, MOLEC, MOPRADEF y MODECULT, ENPECYT, ENH, ENDUTIH y MOCIBA.
 - **DGEGSPyJ**: ENIVPE (muestreo probabilístico).



TNR para cada variable principal: Resultados

- Para los proyectos con **muestreo probabilístico** y con **muestreo no probabilístico**, se observó heterogeneidad en la interpretación de la **definición del denominador** (i.e. número de unidades de observación donde se intentó levantar la variable principal z) e **información no disponible** para realizar el cálculo conforme a la fórmula:
 - **DGEE**. Dado que las variables principales son obligatorias, el denominador se conforma de todas las unidades de observación en la muestra, pues en todos los casos se tiene certeza de que se intentaría obtener la información sobre éstas. En este caso el cálculo para todas las variables principales es el mismo y coincide con la tasa de no respuesta antes de imputación a nivel unidad de observación.
 - **DGES**. Se identificaron 3 variantes de acuerdo a la disponibilidad de información para el cálculo del denominador: a) Se tiene información disponible; b) no se tiene la información porque no todos los cuestionarios son capturados (e.g. cuestionarios en papel); y c) es probable que la información esté en el sistema, pero se requiere hacer una preparación de la información para explotarla (e.g. recolección por medios electrónicos).
 - **DGEGSPyJ**. El denominador podría ser definido de acuerdo a dos bases de datos que se guardan en el sistema: una denominada “Base no validada” y la otra “Base validada”, sin embargo, en ningún caso se asegura que la información disponible corresponda al total de casos en donde se realizó la pregunta.

TNR para cada variable principal: Conclusión

- En los proyectos con **muestreo probabilístico y no probabilístico** no es factible el cálculo estandarizado en el corto plazo para todos los proyectos.
- Se propone elaborar una propuesta de implementación de este indicador para los proyectos con muestreo y presentarla en la primera sesión del 2018, considerando dos opciones:
 1. Recopilar en los sistemas todo el registro histórico (linaje) del dato.
 2. Adecuar la fórmula a la información disponible.
- Se reprograma la presentación de resultados del diagnóstico para los **proyectos censales y registros administrativos** para la primera sesión del 2018.



Publicación de indicadores



Tipos de proyectos de producción de información por grado de madurez*

- **Proyectos experimentales:** son aquellos que **generan información para temas emergentes o coyunturales**. Su continuidad no ha sido determinada.
- **Proyectos recurrentes:** se generan con una **periodicidad determinada** que permite dar seguimiento a las variables de interés **durante tres o más momentos en el tiempo** ya sea de forma continua o periódica.
- **Proyectos recurrentes y estandarizados:** **siguen las fases del Modelo del Proceso Estadístico y Geográfico** a partir de las cuales se incorporan controles de calidad estandarizados.
- **Información de Interés Nacional:** cumplen con **los requisitos que establece la Ley del SNIEG** además de aquellas características que determine la Junta de Gobierno.

* *Tipología definida en el PNEG 2013-2018. Actualización 2017*

Publicación indicadores de precisión estadística

- Para los **proyectos con muestreo probabilístico y muestreo no probabilístico** se propone calcular y reportar de manera estandarizada en los metadatos los indicadores de precisión estadística aprobados por el CAC, conforme a las especificaciones de las fichas técnicas.
- El universo inicial de proyectos se definiría con base en el **Calendario de difusión de información estadística y geográfica y de Interés Nacional del INEGI, 2018**.
 - Se reportarían los indicadores para cada levantamiento de las encuestas publicadas en 2018, independientemente del periodo de referencia de la información.
 - **Las encuestas publicadas en el sitio del INEGI en internet no incluidas en el calendario de difusión también reportarían sus indicadores de precisión estadística en los metadatos.**
 - En el caso de proyectos catalogados como experimentales se calcularían y reportarían los indicadores para uso interno (incluidos proyectos publicados en boletines).
 - En el primer trimestre del 2018 se definirá el formato para reportar retroactivamente los indicadores a partir del segundo trimestre.

Encuestas con muestreo probabilístico registradas en el calendario de difusión 2018 que reportarían el CV (EE e IC) y TNR antes de imputación a nivel unidad de observación.

DGEE		DGES		DGEOSPJ		DGGMA	
Proyecto	Grado de madurez	Proyecto	Grado de madurez	Proyecto	Grado de madurez	Proyecto	Grado de madurez
EMIM*	IIN	ENOE	IIN	ENVIPE	IIN	Módulo de Hogares y Medio Ambiente de la ENH	Recurrente
EMS*	IIN	ENCO	Recurrente	ENVE	IIN		
EMEC*	IIN	MOPRADEF	Recurrente	ENCIG	IIN		
ENEC	IIN	ENH	Recurrente	ENSU	Recurrente		
ENA	Recurrente	MODECUL	Recurrente				
EAEC	Recurrente	MOLEC	Recurrente				
EMOE	Recurrente	BIARE	Recurrente				
EAC*	Recurrente						
EASPNF*	Recurrente						
EAT*	Recurrente						
EAIM*	Recurrente						

*Proyectos con muestreo combinado

Encuestas por muestreo no probabilístico de la DGEE

- Calcularían y reportarían la **Cobertura de la variable de diseño y la TNR antes de imputación a nivel unidad de observación**

Muestreo combinado

Grado de madurez

- | | |
|---|------------|
| 1. Encuesta Mensual de la Industria Manufacturera (EMIM) | IIN |
| 2. Encuesta Mensual sobre Empresas Comerciales (EMEC) | IIN |
| 3. Encuesta Mensual de Servicios (EMS) | IIN |
| 4. Encuesta Anual del Comercio (EAC) | Recurrente |
| 5. Encuesta Anual de Servicios Privados No Financieros (EASPNF) | Recurrente |
| 6. Encuesta Anual de Transportes (EAT) | Recurrente |
| 7. Encuesta Anual de la Industria Manufacturera (EAIM) | Recurrente |



GRACIAS

