

Control de calidad
de la BDD del
Directorio de
Empresas y
Establecimientos
DIEE-2018

Diciembre, 2019



Buenas cifras,
mejores vidas

Contenido

Introducción..... 3

Objetivo 3

Objetivos específicos 3

1. Cálculo de Indicadores 3

1.1. Consideraciones preliminares 3

1.1. Cálculo y selección de la muestra 5

1.2. Resultados 6

1.2.1. Indicadores por cobertura de identificación..... 7

1.2.2. Indicadores por cobertura de identificación..... 10

1.2.3. Indicadores por variables demográficas 14

1.2.4. Indicadores por imputación de las principales variables 15

Introducción

El Directorio de Empresas y Establecimientos (DIEE) es un producto estadístico importante del Instituto Nacional de Estadística y Censos (INEC), ya que, en su función de proveedor oficial de las estadísticas económicas para Ecuador, pone a disposición esta herramienta clave, que proporciona información de las empresas con una cobertura casi completa del sector formal, para obtener estadísticas para la toma de decisiones de política económica.

El DIEE una operación estadística basada en registros administrativos cuyas principales fuentes de información proviene del Instituto Ecuatoriano de Seguridad Social (IESS), del Servicio detentas Internas (SRI) y de diferentes encuestas a empresas que el INEC realiza. Con ello se produce información estadística actualizada, para satisfacer requerimientos sobre marcos muestrales para la mayoría de las encuestas dirigidas a empresas; directorio que se compone con unidades económicas públicas o privadas con personería jurídica propia o sin ella, que se encuentran dentro del territorio nacional y que generaron alguna actividad económica en el periodo de estudio.

Con este antecedente y dado los resultados que se obtuvieron para la publicación del DIEE-2018, el presente documento va enfocado a obtener indicadores de calidad que permitan medir su cobertura y calidad.

Objetivo

Calcular indicadores que permitan medir la cobertura del Directorio de Empresas y Establecimientos que publica la unidad de Gestión de Estadísticas Económicas en Base a Registros Administrativos.

Objetivos específicos

- Calcular indicadores por cobertura de identificación.
- Calcular indicadores por cobertura de estratificación.
- Calcular indicadores por cobertura de imputación.
- Calcular indicadores por variables demográficas.

1. Cálculo de Indicadores

1.1. Consideraciones preliminares

Para el cálculo de los indicadores antes mencionados se utilizará la base de datos completa final (datos para publicación) DIEE-2018, la obtención de los indicadores se ejecutará en dos fases:

1.1.1. Primera Fase

En esta fase se procesa la **BDD completa** y se calcula los siguientes indicadores:

- Porcentaje de empresas del DICE que disponen detalle de dirección completa.
- Porcentaje de empresas que se actualizó detalle de dirección completa.
- Porcentaje de empresas que disponen información de actividad económica a 4 y 6 dígitos de la CIIU.
- Porcentaje de actividad económica a 4 y 6 dígitos de la CIIU que fueron actualizadas.
- Porcentaje de empresas que se actualizó información de ubicación geográfica.
- Porcentaje de empresas que disponen de información de ubicación geográfica.
- Porcentaje de empresas imputadas para definir tamaño por ventas.
- Porcentaje de empresas imputadas para definir tamaño por plazas de empleo registrado y remuneraciones.
- Porcentaje de empresas falsas activas.

Las variables de la BDD esenciales que intervienen en esta fase son: Dirección, Código de actividad económica a 4 y 6 dígitos de la CIIU. Ubicación de geografía, ventas, empleo y remuneraciones.

1.1.2. Segunda Fase

En esta fase se selecciona de la BDD DICE2018 completa, las empresas definidas en los tamaños: **grandes, medianas A y B**, se realiza esta selección dado que no se dispone de tiempo y recursos para analizar la BDD completa, debido a ello de los tamaños más importantes se calcula una muestra, la cual permitirá obtener los siguientes indicadores:

- Porcentaje de empresas que presentaron error de dirección.
- Porcentaje de empresas que presentaron error de codificación.
- Porcentaje de empresas que presentaron error de ubicación geográfica.

Para gestionar la muestra obtenida de los tamaños de empresas antes descritas, se utilizará el sistema Call Center, en el cual se verificará el siguiente listado de variables:

- Identificador de empresa, RUC, razón social y nombre comercial (esenciales para llevar un control de las empresas a revisar).
- Código CIIU.
- Descripción de la Actividad Económica.
- Ubicación Geográfica.
- Variables de Dirección.
- Variables de contacto.

Para ello se establece las siguientes directrices:

- Investigar si la empresa estuvo activa en el año 2018.
- Investigar la actividad económica de la empresa en el año 2018.
- Investigar la ubicación geográfica a nivel de provincia, cantón y su dirección en el año 2018.

Esta actividad será realizada por el equipo verificador de GEERA y lo hará directamente con la empresa a través de llamada telefónica para investigar la información requerida. Únicamente en caso de que no se logre ubicar a la empresa por la vía antes descrita, deberá buscar la información mediante páginas WEB, de

preferencia en la página de la empresa si es que esta la dispone.

1.1. Cálculo y selección de la muestra

La muestra se calcula del universo de empresas obtenidas de la BDD total final, definidas en los tamaños: Grande, medianas A y B; por lo tanto esta será la población objetivo, la muestra a calcular debe reproducir las características del universo o población; por ello, la muestra debe ser representativa de esta, al considerar un tamaño de muestra suficiente, un nivel de confianza seguro y un máximo de error permitido; con estas características definidas se puede argumentar que la muestra obtenida es representativa de la población objetivo.

Para aplicar las características descritas, se utiliza la fórmula para muestreo aleatorio simple:

$$n = \frac{N * Z_{\alpha}^2 * \rho * q}{(N - 1) * e^2 + Z_{\alpha}^2 * \rho * q}$$

Donde:

- n = Tamaño de la muestra
- N = Total poblacional (sumatoria de los tres estratos)
- Z_α = Variables de estandarización dado para el nivel de confianza
(Valor medio para lograr el nivel de confianza deseado, distribución de Gauss)
- ρ = Proporción esperada en el estudio (éxito)
- q = Proporción no esperada en el estudio (fracaso)
- e = Margen de error máximo admitido en el estudio

Para el estimador "p" se considera el porcentaje de 0.5, y para la probabilidad de fracaso de "q" de 0.5; con un nivel de confianza del 95% (según tablas estadísticas) Z_α=1.96, y con el fin de construir indicadores precisos el error "e" será igual a 3,17%.

Bajo lo expuesto se obtuvo una muestra de 906 empresas. Dado que se cuenta con tamaños de empresa, la muestra obtenida se reparte para cada estrato; este tipo de muestreo es una extensión del muestreo aleatorio simple, debido que en primer lugar se calcula el tamaño total de la muestra y posterior a ello y función al número de elementos con el que está compuesto cada estrato se redistribuye la muestra; esto último se lo conoce como afijación proporcional tal como se muestra en la tabla 2. Para ello se utiliza la siguiente formulación:

$$n_i = \left(\frac{N_i}{N} * n \right) * 100$$

Tabla 1 Tamaño de la muestra

Tamaño Empresas	Total, Universo	Muestra	Porcentaje frente al total	Porcentaje acumulado muestra
Mediana empresa "A"	8.529	173	46,01%	46,01%
Mediana empresa "B"	5.749	117	31,12%	77,13%
Grande empresa	4.260	86	22,87%	100,00%
Total, muestra	18.538	376	100%	

Fuente: BDD del DIEE-2018

1.1.1. Tiempo considerado para la verificación de la información en Call Center

El tiempo que utilizará el verificador para realizar esta actividad, se calculó con una piloto inicial, en la cual se determinó que el verificador debe emplear 15 minutos por empresa; tiempo en el que se consideró varios factores negativos como: no todas las empresas cuentan con números de teléfonos fijos correctos; tampoco todas las empresas disponen de página WEB; aun cuando se verifica que si existe información en las fuentes el SRI, SUPERCIAS y diferentes páginas WEB, estas difieren de la información que se registra en la base de datos del DIEE; adicional a ello está la búsqueda de códigos CIUU para actividades económicas muy específicas.

Si se emplea el 100% de tiempo a la actividad de verificará y actualizará la información en el sistema Call Center, en promedio el verificador realizará 32 registros diarios. A continuación, se describe el tiempo a ser utilizado:

Tabla 2 Distribución del tiempo

Tamaño Empresas	Muestra	Carga diaria 100%	Tiempo en días al 100%	% de avance acumulado 100%	Carga diaria 75%	Tiempo en días al 75%	% de avance acumulado 90%
Mediana empresa "A"	173	32	5,41	45,69%	24	7,21	46,01%
Mediana empresa "B"	117	32	3,66	76,27%	24	4,88	77,13%
Grande empresa	86	32	2,69	100,00%	24	3,58	100,00%
Total muestra	376		11,76			15,67	

Fuente: BDD del DIEE-2018

Nota*: Cuando se calcula solo con el 90% de tiempo para la actividad, el verificador debe gestionar 29 empresas, se redondea a 30 ya que realiza otras actividades.

Para realizar la actividad de verificación de la muestra obtenida, se debe utilizar un total de 16 días si el verificador utiliza solo el 75% de su tiempo, ya que el mismo debe realizar otras actividades.

1.2. Resultados

El cálculo de los indicadores del control de calidad de la BDD del DIEE-2018, se realizó bajo cuatro categorías de variables: identificación, estratificación, demografía y otras variables como las imputaciones; a continuación, el detalle de los indicadores obtenidos:

De la muestra obtenida para verificar en el sistema Call Center se comprobó el 95,48%, lo restante (4,52%) corresponde a empresas solo gestionadas que no se logró obtener información por ninguno de los medios establecidos en las directrices; con estos resultados se extrapoló al universo.

Tabla 3 Gestión de la muestra*

Tamaño Empresas	Total Universo	Muestra	Verificadas (terminada)	Porcentaje frente al total
Mediana empresa "A"	8.529	173	163	94,22%
Mediana empresa "B"	5.749	117	111	94,87%
Grande empresa	4.260	86	85	98,84%
Total muestra	18.538	376	359	95,48%

Fuente: Muestra Call Center DICE 2018.
Elaboración: GEERA.

1.2.1. Indicadores por cobertura de identificación

- ✓ Porcentaje de empresas del DICE-2018 que disponen detalle de dirección completa.

Para calcular este indicador se considera las variables que intervienen en la estructura de identificación, tal como se lista en la siguiente tabla:

Tabla 4 Estructura para definir registros con dirección completa.

No. de Bloque	Estructura de Dirección completa	Bloque selección	Condición por bloque	Condición general
1	Calle principal	1	≥ 1	≥ 3
	Intersección	1		
	Numero Kilometro			
2	Urbanización	2,3,4, 5 y 6	≥ 1	≥ 3
	Nombre edificio			
	Número piso			
	Número oficina			
3	Ciudad/ci	2,3 y 4	≥ 1	≥ 3
	Barrio			
	Referencia Manzana			

Fuente: Base de datos DICE 2018.
Elaboración: GEERA.

Aquellos registros que no cumplieron la condición general se determinaron como direcciones no válidas, los bloques de selección se establecen en función de:

- Si no existe ningún dato en los campos de dirección (calle principal y secundaria).

- b) Si solo existe 1 dato en cualquiera de los campos: número, kilómetro, urbanización, nombre edificio, número piso y número oficina.
- c) Si solo existe un dato en los campos: ciudadela, barrio, referencia y manzana

Si no se cumple la condición en cada uno de los bloques se determinó como dirección no válida. Para calcular cada uno de los bloques se utilizó el programa SPSS con la función recodificar variables en nuevas variables con ello se calculará las nuevas variables. Posterior a ello se empleó la siguiente formulación, la cual se aplicó tanto al universo de la población como a los estratos por tamaño de empresas:

$$PEDC_t = \frac{TESDC_t}{TEA_t} * 100$$

Donde:

PEDC_t= Porcentaje de empresas con dirección completa en el año de referencia.

TESDC_t= Total empresas sin dirección completa en el año de referencia.

TEA_t= Total empresas activas en el año de referencia.

Resultados:

$$PEDC_t = \frac{776.298}{899.208} * 100$$

$$PEDC_t = 86,33\%$$

Se puede evidenciar que el 86,33% de empresas disponen de dirección para su ubicación tal como se muestra en la siguiente tabla.

Tabla 5 total empresas que disponen o no de información de dirección.

Detalle de dirección del bloque 1	# de empresas	%
Dirección vacía	12.184	1,35%
Dirección en al menos dos de las variables	110.726	12,31%
Dirección completa	776.298	86,33%
Total	899.208	100%

Fuente: Base de datos DICE 2018.

Elaboración: GEERA.

✓ Porcentaje de empresas que se actualizó detalle de dirección.

Para el cálculo de este indicador se realizó un conteo de todas las empresas que se actualizaron dirección completa (registro_fecha_direccion), en función de las diferentes fuentes como: IND=Industriales, CALL=Cal Center, SRI= Servicio de Rentas Internas, entre otras y que correspondan al año de referencia. Se aplicó la siguiente fórmula tanto para el universo como para los estratos de empresas.

$$PEAD = \frac{TESAD}{TEA} * 100$$

Donde:

PEAD= Porcentaje de empresas que se actualizaron dirección completa en el año de referencia.

TESAD= Total empresas sin actualizar dirección completa en el año de referencia.

TEA= Total empresas activas en el año de referencia.

Resultados:

$$PEAD = \frac{191.295}{899.208} * 100$$

$$PEAD = 21,27\%$$

La base del DICE 2018 presenta una actualización de la variable dirección del 21,27% del directorio total. De este total general se analiza a las empresas grandes, medianas A y B, con lo cual se puede observar que: este grupo de empresas se actualizó en un 57,88%, de ello, el mayor porcentaje de actualización de detalle de dirección corresponde a las empresas Mediana B (66,97%), tal como se muestra en la siguiente tabla:

Tabla 6 Actualización de detalle de dirección, según tamaño de empresa.

Actualización de detalle de dirección	Empresas Grandes	%	Empresas Mediana B	%	Empresas Mediana A	%	Total	%
SI	1.369	32,14%	3.850	66,97%	5.510	64,60%	10.729	57,88%
NO	2.891	67,86%	1.899	33,03%	3.019	35,40%	7.809	42,12%
Total	4.260	100%	5.749	100%	8.529	100%	18.538	100,00%

Fuente: Base de datos DICE 2018.

Elaboración: GEERA.

✓ **Qué porcentaje de empresas presentan error en la dirección:**

Resultados:

Aproximadamente 1.346 empresas presentaron error en la dirección, tal como se observa a continuación:

Tabla 7 Porcentaje de empresas que presentaron error de dirección.

Error dirección	Total empresa	%
SI	1.346	7,26%
NO	17.192	92,74%
Total	18.538	100%

Fuente: Base de datos DICE 2018.

Elaboración: GEERA.

De las empresas verificadas el 92,74% se ubica en la misma dirección, mientras que el 7,26% de las empresas registra una ubicación diferente al de la base de datos DICE-2018.

Ahora bien, analizando el indicador por tamaño de empresa resulta que el porcentaje obtenido se distribuye uniformemente en cada uno de los estratos, y se puede observar que las empresas "Mediana A" son las que más error de dirección presentan (8,59%), tal como se observa en la siguiente tabla:

Tabla 8 Distribución de empresas que presentaron error de dirección por tamaño de empresa.

Error dirección	Empresas Grandes	%	Empresas Mediana B	%	Empresas Mediana A	%
SI	251	5,88%	363	6,31%	733	8,59%
NO	4.009	94,12%	5.386	93,69%	7.796	91,41%
Total	4.260	100%	5.749	100%	8.529	100%

Fuente: Base de datos DIEE 2018.

Elaboración: GEERA.

1.2.2. Indicadores por cobertura de identificación

El cálculo se lo realiza tanto para la población como para los estratos de estudio.

✓ Porcentaje de empresas que disponen información de actividad económica a 4 y 6 dígitos de la CIIU Rev. 4.0.

$$PAE_i = \frac{TESAE_t}{TEA_t} * 100$$

Donde:

PAE_i= Porcentaje de empresas que disponen código de actividad económica de la CIIU 4.0 a nivel de 6 dígitos.

TESAE_t= Total empresas sin actividad económica en el año de estudio.

TEA_t= Total empresas activas en el año de estudio.

Resultados:

El 100% de las actividades registradas en la BDD del DIEE-2018 dispone de códigos a nivel de 4 dígitos de la CIIU Rev. 4.0.

Para los códigos a nivel de 6 dígitos de la CIIU Rev. 4.0 se indica que:

$$PAE = \frac{758}{899.208} * 100$$

$$PAE = 0,08\%$$

El 0,08% de empresas no disponen de código de actividad a 6 dígitos de la CIIU Rev. 4.0 (incluye todos los estratos de empresas). Si separamos las empresas que no dispone de código de actividad económica en: grupo 1 empresas Grandes y medianas A y B y grupo 2 las demás (Pequeña empresa y microempresa), se puede verificar que el primer grupo aporta apenas con el 0,13% y las demás con el 99,87%.

Tabla 9 Empresas por tamaño que no dispone de actividad económica a 6 dígitos de CIIU Rev. 4.0

Tamaño de empresas	# de empresas	%
Grupo 1 (Empresas grandes, mediana A y B)	1	0,13%
Grupo 2 (Pequeña empresa y microempresa)	757	99,87%
Total	758	100%

Fuente: Base de datos DIEE 2018.
Elaboración: GEERA.

Ahora bien, de este porcentaje (0.13%) que no disponen código de actividad, lo representa las empresas del estrato "Empresas Mediana A", tal como se muestra en la siguiente tabla:

Tabla 10 Distribución de empresas que no disponen de código de actividad económica a 6 dígitos de la CIIU Rev. 4.0 por tamaño de empresa.

Dispone de actividad económica	Empresas Grandes	%	Empresas Mediana B	%	Empresas Mediana A	%	Total	%
No	0	0,00%	1	0,02%	0	0,00%	1	0,01%
Si	4.260	100,00%	5.748	99,98%	8.529	100,00%	18.537	99,99%
Total	4.260	100%	5.749	100%	8.529	100%	18.538	100%

Fuente: Base de datos DIEE 2018.
Elaboración: GEERA.

✓ **Porcentaje de actividades económicas de empresas a nivel de 4 y 6 dígitos de la CIIU Rev. 4.0 que fueron actualizadas.**

Para calcular este indicador se toma la variable "fecha_registro_actividad", cuyos registros tengan año de actualización 2019.

$$PAE_t = \frac{TEAE_t}{TEA_t} * 100$$

Donde:

PAE_t= Porcentaje de empresas que se actualizó actividad económica a nivel de 4 y 6 dígitos de la CIIU 4.0.

TEAE_t= Total empresas que se actualizó actividad económica a nivel de 4 y 6 dígitos de la CIIU 4.0.

TEA_t= Total empresas activas en el año de estudio.

Resultados:

$$PAE = \frac{148.869}{899.208} * 100$$

$$PAE = 16,56\%$$

Para el año de referencia el Directorio de Empresas y Establecimientos actualizó la variable actividad económica en un 16,56%; de ello a continuación se selecciona a las empresas definidas en el estrato de empresas: grandes, mediana A y B; se puede

observar que el mayor porcentaje de actualización lo presentan las empresas "mediana A" con un 10,12%.

Tabla 11 Distribución de empresas por tamaño que se actualizó código de actividad económica a nivel de 6 dígitos de la CIIU 4.0.

Se actualizó Actividad Económica	Empresas Grandes	%	Empresas Mediana B	%	Empresas Mediana A	%	Total	%
SI	224	5,26%	519	9,03%	863	10,12%	1606	8,66%
No	4.036	94,74%	5.230	90,97%	7.666	89,88%	16.932	91,34%
Total	4.260	100%	5.749	100%	8.529	100%	18.538	100%

Fuente: Base de datos DICE 2018.

Elaboración: GEERA.

✓ **Porcentaje de empresas que presentaron error de codificación***

Para calcular este indicador se toma la variable cambio de sector económico de los registros verificados en el sistema Call Center.

Según la verificación realizada, se tiene que 673 empresas presentaron error en la codificación de la actividad económica, que representa el 3,63%, tal como se muestra a continuación:

Tabla 12 Porcentaje de empresas que presentaron error de codificación.

Error codificación	Total empresa	%
SI	673	3,63%
NO	17.865	96,37%
Total	18.538	100%

A continuación se analiza el error de codificación solo de los siguientes estratos de empresas:

Tabla 13 Porcentaje de empresas que presentaron error de codificación Actividad Económica.

Error codificación	Empresas Grandes	%	Empresas Mediana B	%	Empresas Mediana A	%
SI	150	3,53%	52	0,90%	471	5,52%
NO	4.110	96,47%	5.697	99,10%	8.058	94,48%
Total	4.260	100%	5.749	100%	8.529	100%

Se puede advertir que las empresas con mayor error de codificación, son las empresas definidas como: empresas "Mediana A" con el 5,52%.

✓ **Porcentaje de empresas que dispone de información de ubicación geográfica.**

Para este indicador se toma las variables "codigo_parroquia", con ello se puede verificar que todos los registros dispongan de geografía y se aplica la siguiente formulación:

$$PEUG_t = \frac{TESUG_{t*}}{TEA_t} \cdot 100$$

Donde:

PEUG= Porcentaje de empresas que disponen ubicación geográfica en el periodo de referencia.

TESUG= Total empresas sin ubicación geográfica en el periodo de referencia.

TEA= Total empresas activas en el periodo de referencia.

Resultados:

$$PEUG = \frac{899.208}{899.208} * 100$$

$$PEUG = 100\%$$

Se puede verificar que el 100% de los registros disponen de ubicación de geografía. Debido a ello los estratos de empresas grande, mediana A y B también disponen de ubicación de geografía al 100%.

✓ Porcentaje de empresas que se actualizó información de ubicación geográfica.

$$PEAUG_t = \frac{TEAUG_t}{TEA_t} * 100$$

Donde:

PEAUG_t= Porcentaje de empresas que se actualizó ubicación geográfica en el periodo de referencia.

TEAUG_t= Total empresas que se actualizó ubicación geográfica en el periodo de referencia.

TEA_t= Total empresas activas en el periodo de referencia.

Resultados:

$$PEAUG = \frac{250.203}{899.208} * 100$$

$$PEAUG = 27,82\%$$

El 27,82% de registros de empresas se actualizó información de ubicación geográfica; este resultado se obtiene a partir de la variable "registro_fecha_geografia".

Seleccionando sólo las empresas grandes, medianas A y B, se verifica que en la base de datos de empresas del DICE 2018, se actualizó información de ubicación geográfica en un 55,21% y el mayor porcentaje lo representa las empresas grandes 77,56%.

Tabla 14. Empresas por tamaño que se actualizó o no información de ubicación geográfica.

Actualización Ubicación Geográfica	Empresas Grandes	%	Empresas Mediana B	%	Empresas Mediana A	%	Total	%
SI	3304	77,56%	2922	50,83%	4009	47,00%	10235	55,21%

No	956	22,44%	2.827	49,17%	4.520	53,00%	8.303	44,79%
Total	4.260	100%	5.749	100%	8.529	100%	18.538	100%

Fuente: Base de datos DICE 2018.

Elaboración: GEERA.

✓ **Porcentaje de empresas presentan error en la ubicación geográfica*.**

Luego de la verificación en el CALL CENTER, se tiene que: El 2,53% de empresas presentaron error en la ubicación geográfica lo que representa 469 registros.

Tabla 15. Empresas por tamaño que presentaron error de ubicación geográfica.

Error geografía	Total empresa	%
SI	469	2,53%
NO	18.069	97,47%
Total	18.538	100%

Fuente: Base de datos DICE 2018.

Elaboración: GEERA.

De las empresas verificadas aproximadamente el mayor porcentaje de error lo representa las empresas Mediana A con el 3,68% y las empresas grandes no presentaron registros con error de ubicación geográfica, tal como se muestra en la siguiente tabla:

Tabla 16 Porcentaje de empresas por tamaño que presentaron error en ubicación geográfica.

Error geografía	Empresas Grandes	%	Empresas Mediana B	%	Empresas Mediana A	%
SI	0	0,00%	155	2,70%	314	3,68%
NO	4.260	100,00%	5.594	97,30%	8.215	96,32%
Total	4.260	100%	5.749	100%	8.529	100%

Fuente: Base de datos DICE 2018.

Elaboración: GEERA.

1.2.3. Indicadores por variables demográficas

Con la revisión de las empresas en Call Center se podrá calcular el porcentaje de empresas que en el año de referencia eran empresas falsas activas de cada uno de los estratos en estudio.

✓ **Porcentaje de empresas falsas activas.**

Las empresas falsas activas hacen referencia a empresas que no registran ventas, empleo o pertenecen al RISE y en el año de referencia se encuentran activas.

Se puede verificar que a nivel de estratos no se presentan registros que contengan con las categorías antes descritas, a continuación, tabla de desglose:

Tabla 17 Porcentaje de empresas falsas activas por tamaño.

Situación 2018 (falsas activas)	Empresas Grandes	%	Empresas Mediana B	%	Empresas Mediana A	%	Total	%
SI	0	0,00%	0	0,00%	0	0,00%	0	0,00%
NO	4.260	100,00%	5.749	100,00%	8.529	100,00%	18.538	100,00%
Total	4.260	100%	5.749	100%	8.529	100%	18.538	100%

Fuente: Base de datos DICE 2018.

Elaboración: GEERA.

1.2.4. Indicadores por imputación de las principales variables

Las imputaciones son procedimientos que permiten asignar información de variables que por algún motivo no se dispone en el período de análisis. Para el caso del DICE, al ser una operación estadística basada en registros administrativos, se realiza imputaciones a aquellas empresas que no disponen datos de ventas para el año de publicación, debido a ello se considera datos de años anteriores, se tomó información del INAR, considerando el valor mensual y obteniendo la suma total para el año. Adicional a ello también se utiliza la información de la encuesta industriales, del SIPRO y las que deriven del análisis del equipo técnico de la DICE, en ello están las siguientes empresas: "EMPRESA PUBLICA DE EXPLORACION Y EXPLOTACION DE HIDROCARBUROS PETROAMAZONAS EP" y "EMPRESA PUBLICA DE HIDROCARBUROS DEL ECUADOR EP PETROECUADOR".

✓ Porcentaje de empresas imputadas para definir tamaño por ventas

$$PIM_t = \frac{TRIM_t}{TEA_t} * 100$$

Donde:

PIM_t = Porcentaje de registros imputados en el periodo de referencia.

$TRIM_t$ = Total de registros imputados por ventas en el periodo de referencia.

TEA_t = Total registros activos en el periodo de referencia.

Resultados:

$$PIM = \frac{2}{899.208} * 100$$

$$PIM = 0.00\%$$

Para el DICE_2018 se imputo un minúsculo número de empresas para definir tamaño de empresas por ventas

✓ **Porcentaje de empresas imputadas para definir tamaño por empleo y remuneración**

Para el caso de las variables empleo y remuneraciones se imputan aquellas empresas que corresponden a Educación y Administración Pública.

Para ello se considera a cada una de las Direcciones Distritales con las Instituciones que se encuentran adscritas a cada una de ellas, procediendo a agruparlas para identificar el empleo mensual que presentaron las instituciones educativas hasta el año 2014 para relacionarlo con el empleo mensual que muestra la Dirección Distrital que la contiene. A partir del año 2015 la cantidad de empleo que muestra la Dirección Distrital se distribuye para cada una de las instituciones que se encuentran a su cargo. Para ello se coloca como valor de empleo, el último que declaró cada institución en el 2014 y se resta la cantidad de empleados que tiene la Dirección Distrital.

Por otro lado, la imputación de remuneraciones se sigue la misma metodología que la de empleo, que para este caso es el valor total de salarios que está asignado a la Dirección Distrital y se distribuye para las instituciones educativas adscritas a la misma.

En función a ello se obtiene los registros de empresas en los cuales se cumple las imputaciones descritas y se aplicará la siguiente formulación:

$$PIM = \frac{TRIM}{TEA} * 100$$

Donde:

PIM= Porcentaje de registros imputados en el periodo de referencia.

TRIM= Total de registros imputados por plazas y/o empleo registrado y remuneraciones en el periodo de referencia.

TEA= Total registros activos en el periodo de referencia.

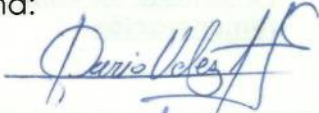
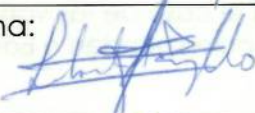

Resultados:

$$PIM = \frac{13.357}{899.208} * 100$$

$$PIM = 1.49\%$$

El porcentaje de empresas imputadas para definir tamaño de las empresas por plazas de empleo registrado, empleo registrado y remuneraciones es del 1.49%, que representa un total de 13.357 empresas.

NOTA: Los resultados que constan con un * se calcularán una vez concluida la etapa de verificación de información el sistema Call Center. El cual se encuentra solo con una persona validando.

Aprobado por:	Darío Vélez	Firma: 
Revisado por:	Libertad Trujillo	Firma: 
Elaborado por:	Rocío Andino	Firma: 
Fecha de aprobación:		26/12/2019
Fecha de actualización		08/01/2020