

Informe de resultados malla I y matrices de actualización

Febrero, 2022.

01.

Informe de resultados malla I y matrices de actualización

Antecedentes

La unidad de gestión de estadísticas económicas en base a registros administrativos y en función al calendario establecido anualmente, calcula y publica los resultados del Directorio de Empresas y Establecimientos DICE, operación estadística que se construye a partir de Registros administrativos, que se obtienen de dos entidades técnicas, autónomas, con personería jurídica y derecho público, como lo son: el Servicio de Rentas Internas (SRI) y el Instituto Ecuatoriano de Seguridad Social (IESS).

En la construcción del DICE, se toma en cuenta estándares mínimos necesarios de acuerdo a experiencias y recomendaciones internacionales, desde su integración hasta llegar a su difusión, etapa que también cuenta con normas de publicación; adicional a ello también se considera las necesidades de usuarios de la operación estadística.

En GEERA se aplican una serie de validaciones que permiten detectar errores en la información de la base de datos las mismas se realizan en el programa R mediante un script que se ajusta a las necesidades del Directorio de empresas; esta serie de validaciones son realizadas por el personal técnico de la unidad, con ello se corrige errores antes de obtener la base de datos final. Además, se realizó una revisión de las matrices de actualización.

Por ello el presente documento constituye una relatoría de los resultados obtenidos al ejecutar el script en R de la Malla I, con las validaciones aplicadas en la base de datos del Directorio de Empresas y Establecimientos DICE 2020 como insumo para la generación de factores de expansión de ENESEM, y de las actualizaciones realizadas.

Desarrollo

- ✓ **Base que se entrega desde el Directorio de Empresas y Establecimientos como insumo para la generación de factores de expansión de ENESEM con fecha 14/02/2022**

El script "Malla_I_v7" está constituida por 115 validaciones de las cuales los resultados se exportan en un documento con formato Excel "Resultados", con las novedades o errores que se encontraron en la base de datos.

- Se transformó la base de datos entregada en txt a Spss, con el nombre de "20220214_empresas_marco_2012_2020", consta de 7'932.864 registros.
- Se ejecutó el script "Malla_I_v7"

Se realizó la revisión de que se hayan realizado las actualizaciones solicitadas en las matrices, se revisaron las siguientes matrices:

- 20220131_Matriz_actualizacion_actividad_economica
- 20220131_Matriz_actualizacion_ventas
- 20220131_Matriz_actualizacion_novedades

✓ Resultados de la base 14/02/2022

La malla se corrió en varias ocasiones debido a los errores que se encontró, a continuación el detalle:

- La primera vez que se corrió se encontró las siguientes novedades

Novedad	Total Registros	Descripción
error_plazas_prom	270.326	Error en sumatoria plazas promedio Al sumar plazas hombres más plazas mujeres promedio no da el total de plazas existe diferencias
Error_empleo_prom	270.623	Error en sumatoria empleo promedio Al sumar empleo hombres más empleo mujeres promedio no da el total de empleo existe diferencias
Cod_fte_ven	327.776	Empresas con fuente de ventas vacío y ventas totales con información
Cod_fte_fecha	131	Empresas sin fecha fuente y con información en fuente ventas
Cod_fte_prov	898.235	Empresas con fuente geografía y sin información en fecha
Cod_prov_fte	989.621	Empresas con provincia y sin fuentes geografía

- Se corrió luego de que los errores fueron corregidos pero se obtuvieron los siguientes resultados:

Novedad	Total Registros	Descripción
error_situacion	1.048.574	Empresas que deberían estar activas no tienen situación en la base
error_plazas_prom	270.326	Error en sumatoria plazas promedio Al sumar plazas hombres más plazas mujeres promedio no da el total de plazas existe diferencias
Error_empleo_prom	270.623	Error en sumatoria empleo promedio Al sumar empleo hombres más empleo mujeres promedio no da el total de empleo existe diferencias

- Al correr por tercera vez se encontró errores en error plazas promedio, se obtuvieron los siguientes resultados:

Novedad	Total Registros	Descripción
Error_plazas_prom	270.389	Error en sumatoria plazas promedio Al sumar plazas hombres más plazas mujeres promedio no da el total de plazas existe diferencias

- Al correr por cuarta y quinta vez se corrigieron los errores antes mencionados tanto de situación como plazas de empleo promedio y empleo promedio.

En las matrices de actualización se encontraron las siguientes novedades:

Variable	Nombre de la matriz	Registros	Actualizado	Novedades
Actividad económica	20220131_Matriz_actualizacion_actividad_economica	219	Si	Un registro no se encuentra en la base pero es porque se registró una absorción
Ventas	20220131_Matriz_actualizacion_ventas	311	Si	Un registro no cuadra ventas porque se registra un cambio de ruc
Novedades	20220131_Matriz_actualizacion_novedades	30	Si	Todos los registros se encuentran actualizados

Novedades:

Se registró una novedad adicional de la malla I, al correr el análisis descriptivo se encontró un error con plazas de empleo último se da por un error en el cálculo de la variable el cuál se describe y justifica a continuación:

“Se identificó un error en el script para el cálculo de último plazas y último empleo que se ejecutó exclusivamente para el año 2020 de la publicación, ya que realicé una simplificación del script que existía antes, al realizar este proceso incluí en una agregación una variable que no había que incluir... este es el script y su error:

```
(select id_empresa_unificado,plazas_total,max(mes) mes from  
diemp.f_empresa_afiliados_mensual where id_periodo_proceso=20 group by 1,2)
```

No se debió haber considerado en la agrupación el valor de plazas, ya que esto hace que no se obtenga el máximo mes por empresa, sino que se obtiene el máximo mes agrupado por empresa y por valor de plazas"

Soluciones:

- En la base de datos del 14/02/2022, se reportó los errores los principales fueron solventados queda pendiente solventar la novedad de empresas con fuente geografía y sin información en fecha de los últimos años 2020.
- Las matrices de actualización se encuentran actualizadas a excepción de los registros reportados por cambios de ruc o absorciones.
- La novedades de plazas último fue por un error en el cálculo que se encuentra justificado.

Conclusión:

- Al final se obtiene una base de datos verificada y validada, los errores fueron analizados y corregidos oportunamente.

Elaborado por:	Jacqueline Imbaquingo	
Revisado por:	Libertad Trujillo	
Fecha:	17/02/2022	

