

ERAMO. Prueba Piloto Aguascalientes

Bitácora de procesamiento

22 de octubre de 2019

Objetivo

El objetivo de esta bitácora es llevar un registro preciso y puntual de cada una de las acciones realizadas para la preparación, validación e integración de la Base de Datos, desde la salida de captura, hasta la versión final, para su explotación y análisis.

1. Descarga de la Base de Datos

La base de datos fue descargada de la plataforma de *Survey Solutions* en formato SPSS (.sav).

Debido a que no se cuenta con licencia de dicho software comercial, se convirtieron las tablas a formato .csv, empleando la aplicación *Stat/Transfer* (software libre).

Una vez descargadas las tablas, el procesamiento de las mismas se realizó completamente en el lenguaje de programación R, excepto en todas aquellas situaciones previas que se describen a continuación, en las que fue necesario hacer ligeros ajustes manuales a ciertos registros de la tabla *municipal.csv* (renombrada a *establecimientos.csv*).

Descripción de la Base de Datos descargada

La carpeta descargada contiene las siguientes tablas (previamente transformadas a formato delimitado por comas .csv):

1. **assignment_actions.csv**: Tabla donde se registran cambios de asignación. Sin registros en la prueba, al no haberse realizado reasignaciones.
2. **interview_actions.csv**: Contiene 2223 registros de cada una de las acciones de cada figura operativa, dentro de la plataforma. Funciona como un registro o historial de intervenciones, para control.
3. **interview_errors.csv**: Tabla donde se registran los errores que la propia plataforma reporta como error de la aplicación. Se tiene un sólo registro, lo que demuestra que sólo en una ocasión la plataforma falló durante el operativo.
4. **interview_comments.csv**: Contiene 105 registros con los comentarios registrados por el entrevistador hacia el supervisor, o bien los realizados por el administrador en la plataforma central (sede).
5. **interview_diagnostics.csv**: 151 registros correspondientes principalmente a los tiempos de duración de las entrevistas.
6. **municipal.csv**: Contiene 151 registros. Muestra los datos generales del establecimiento, coordenadas, control de la entrevista y campo de observaciones.
7. **lista_residuos.csv**: Tabla con 157 registros, que corresponden al tipo y número de residuos declarados por los establecimientos.
8. **lista_tiporesiduos.csv**: Con 199 registros, comprende la captación de las características, cantidades y destinos de los residuos declarados en los establecimientos.

Descripción de cambios aplicados

- 1) Tabla *assignment_actions.csv*, no se toma en cuenta para conformar la BD de explotación, por no contener registros.
- 2) Tabla *interview_errors.csv*, contiene sólo un registro. Tampoco se toma en cuenta para conformar la BD de explotación, por tratarse de información no relevante para el procesamiento.
- 3) Tabla *interview_actions.csv*, contiene todos los registros de movimientos de las diversas figuras operativas de la encuesta. Tampoco se toma en cuenta para conformar la BD de explotación, por tratarse de información no relevante para el procesamiento.
- 4) Se renombra la tabla *municipal.csv* a *establecimientos.csv*, se crea una copia de la misma y se nombra *establecimientos1.csv*. Sobre esta copia se realizan todas las modificaciones subsiguientes, a efecto de poder realizar cambios manuales en la misma. Todas las tablas revisadas y validadas que se generan según los procedimientos descritos en la presente bitácora, conformarán la base de datos para explotación, y se denominarán agregando el número 2 al nombre original de la tabla, a efecto de diferenciarlas como segunda versión de las mismas.
- 5) Para los comercios ambulantes, en los que se asignó durante el operativo un ID (campo V4) con un formato *número_de_equipo-consecutivo_del_establecimiento_ambulante*, se reasigna un ID numérico, concatenando los siguientes números: *99999 + Número de equipo durante el operativo + Consecutivo del establecimiento ambulante*, a efecto de homogenizar el campo a tipo numérico, y filtrar con mayor facilidad dichos establecimientos en la explotación.
- 6) Se agregan a 30 registros en la tabla *establecimientos1.csv*, los datos de latitud y longitud. Este proceso debió llevarse a cabo de forma manual, debido a las dificultades para ubicar cada unidad económica en el mapa. Quedaron algunas unidades económicas sin asignar coordenadas, debido a que no se les pudo ubicar, sin embargo, se trata de establecimientos no permanentes.

Las coordenadas asignadas fueron:

Llave	ID (DENUE)	Latitud	Longitud
03-05-58-94	28013	21.87301	-102.2584
04-05-29-04	6790694	21.88275	-102.2583
07-42-68-66	36324	21.88395	-102.2971
12-54-30-22	52082	21.88918	-102.2517
17-23-45-49	6282263	21.87438	-102.2553
18-90-43-80	48987	22.07548	-102.2734
20-39-37-73	29226	21.87352	-102.2464
22-93-40-85	7565	21.89292	-102.2750
23-97-42-63	1819	21.89342	-102.3156
31-01-70-00	8221	21.89126	-102.2725
32-78-66-91	25930	22.07807	-102.2727
37-67-74-84	51662	21.88889	-102.2739
39-20-84-48	99999528	21.87890	-102.2929
41-05-66-05	51653	21.88925	-102.2739
41-72-60-47	6886730	21.79579	-102.7152
47-44-56-58	8292	21.88743	-102.2726
48-40-47-85	6282322	21.89054	-102.2739
52-81-44-27	51657	21.88925	-102.2736
53-98-17-94	6758020	22.09442	-102.2964
59-86-22-78	25962	22.07687	-102.2730
60-62-73-58	99999540	21.87863	-102.3124
65-31-27-49	27688	21.89126	-102.2725
67-78-81-26	49755	20.53786	-102.3842

Llave	ID (DENUE)	Latitud	Longitud
69-25-05-13	10355	21.87609	-102.2546
70-15-01-95	8295	21.88676	-102.2728
72-22-26-38	5175	21.87007	-102.2954
75-96-72-76	6166248	22.16778	-102.3456
78-93-50-99	8249	21.89325	-102.2707
83-39-13-04	8225	21.89126	-102.2725
84-05-91-99	6758020	22.09415	-102.2962

- 7) En un establecimiento se detectó que las coordenadas no eran correctas (tabla *establecimientos1.csv*), por lo que se corrigieron en la BD de forma manual:

Llave	ID (DENUE)	Latitud original	Longitud original	Latitud corregida	Longitud original
32-04-68-91	8974	21.8477653	-102.2808157	21.832729	-102.289223

- 8) Se detectaron también fechas incorrectas (año 2000), provenientes de alguno de los equipos empleados en la captura (tabla *establecimientos1.csv*), cuyo sistema operativo estaba mal configurado. Se blanquearon 4 datos de fecha debido a la imposibilidad de recuperarlos, correspondientes a las llaves siguientes: 54-31-39-72, 55-95-73-76, 58-57-15-91.
- 9) Se modificaron en la tabla *establecimientos1.csv*, de forma manual los códigos del campo V3 (que define si es prueba de configuración o levantamiento) del valor 1 al 2, para las siguientes llaves: 53-27-23-59, 98-38-33-96, 77-06-67-30.
- 10) Se cambia manualmente en la tabla *establecimientos1.csv* el valor del campo V4 (que corresponde al ID del DENUE) en los siguientes registros, que se detectaron erróneos:

Llave	V4 erróneo	V4 corregida
66-96-44-94	5000	6280911
65-51-61-06	2168	1168
22-07-32-82	1198	1249
96-57-45-57	3402	34012
24-07-91-89	143	99999143

- 11) Se eliminan los siguientes campos para cada tabla, sea por tratarse de campos sin información (como en el caso de campos reservados a captura de texto que no fue necesario utilizar), o por contener información de control del propio sistema, como son el registro de errores y asignaciones, que en realidad no resultan de utilidad para la explotación:

Tabla	Campo eliminado	Observaciones
establecimientos1.csv	V14	Campo para código de reprogramación. Se utilizó para casos que serán filtrados como entrevistas sin respuesta, por lo que es innecesario mantenerlo
establecimientos1.csv	V15	Campo para nombre de persona que atendería en la reprogramación. Se utilizó para casos que serán filtrados como entrevistas sin respuesta, por lo que es innecesario mantenerlo

Tabla	Campo eliminado	Observaciones
establecimientos1.csv	V16	Campo para fecha de reprogramación. Se utilizó para casos que serán filtrados como entrevistas sin respuesta, por lo que es innecesario mantenerlo
establecimientos1.csv	V26_1	Campo adicional para texto. No se utilizó
establecimientos1.csv	V26_2	Campo adicional para texto. No se utilizó
establecimientos1.csv	V26_3	Campo adicional para texto. No se utilizó
establecimientos1.csv	V26_4	Campo adicional para texto. No se utilizó
establecimientos1.csv	V27_0	Campo adicional para texto. No se utilizó
establecimientos1.csv	V30	Campo del sistema. Sin datos
establecimientos1.csv	V31	Campo del sistema. Datos sin utilidad para la explotación
establecimientos1.csv	V32	Campo del sistema. Datos sin utilidad para la explotación
establecimientos1.csv	V33	Campo del sistema. Datos sin utilidad para la explotación
establecimientos1.csv	V34	Campo del sistema. Datos sin utilidad para la explotación
interview__diagnostics.csv	V3	Campo del sistema. Datos sin utilidad para la explotación
interview__diagnostics.csv	V5	Campo del sistema. Datos sin utilidad para la explotación
interview__diagnostics.csv	V6	Campo del sistema. Datos sin utilidad para la explotación
interview__diagnostics.csv	V7	Campo del sistema. Datos sin utilidad para la explotación
interview__diagnostics.csv	V8	Campo del sistema. Datos sin utilidad para la explotación
interview__diagnostics.csv	V9	Campo del sistema. Datos sin utilidad para la explotación
interview__comments.csv	V3	Campo del sistema. Datos sin utilidad para la explotación
interview__comments.csv	V4	Campo del sistema. Datos sin utilidad para la explotación
interview__comments.csv	V5	Campo del sistema. Datos sin utilidad para la explotación
interview__comments.csv	V7	Campo del sistema. Datos sin utilidad para la explotación
interview__comments.csv	V11	Campo del sistema. Datos sin utilidad para la explotación
interview__comments.csv	V12_0	Campo adicional para texto. No se utilizó
interview__comments.csv	V12_1	Campo adicional para texto. No se utilizó
lista_tiporesiduos.csv	V5_0	Campo adicional para texto. No se utilizó

Tabla	Campo eliminado	Observaciones
lista_tiporesiduos.csv	V5_1	Campo adicional para texto. No se utilizó
lista_tiporesiduos.csv	V14_0	Campo adicional para texto. No se utilizó
lista_tiporesiduos.csv	V14_1	Campo adicional para texto. No se utilizó
lista_tiporesiduos.csv	V18	Campo para especificar otra unidad para residuos a bancos de alimentos. No se utilizó

12) Se elimina el registro con llave 99-53-17-80 de la tabla *establecimientos1.csv*, por no contener ninguna información.

13) Se asignan nombres a las variables de las tablas resultantes, de acuerdo con el FD previamente diseñado, conforme a lo siguiente:

Tabla: *establecimientos1.csv*

Campo	Mnemónico asignado	Descripción
V1	llave	Llave de entrevista asignada por el sistema
V2	folio_sist	Folio asignado al registro por el sistema
V3	tipo	Tipo de entrevista
V4	id_denue	Identificador del establecimiento según su registro en el DENU
V5	latitud	Latitud de la coordenada geográfica
V6	longitud	Longitud de la coordenada geográfica
V7	prec	Precisión de la coordenada geográfica
V8	altitud	Altitud SNM de la coordenada geográfica
V9	hora_cg	Hora de registro de la coordenada geográfica
V10	planeado	Define si el establecimiento corresponde a lo planeado en muestra
V11	nvo_establec	Nombre del nuevo establecimiento
V12	saludo	Define si se localiza al dueño o encargado
V13	acepta	Define si el informante acepta la entrevista al momento
V17	consumo	Tipo de consumo de los alimentos que se manejan
V18	mercado	Tipo de mercado de los alimentos que se manejan
V19	residuos_vc	Existencia de Residuos de vegetal crudo

Campo	Mnemónico asignado	Descripción
V20	residuos_vp	Existencia de Residuos de vegetal procesado
V21	residuos_ac	Existencia de Residuos de animal crudo
V22	residuos_ap	Existencia de Residuos de animal procesado
V23	residuos_ot	Existencia de Residuos de otro tipo
V24	ot_residuos_e	Especifica residuos de otro tipo
V25	despedida	Se leyó despedida
V26	obs_1	Observaciones del entrevistador
V26_0	obs_2	Observaciones del entrevistador
V27	obs_3	Observaciones del entrevistador
V28	obs_4	Observaciones del entrevistador
V29	obs_5	Observaciones del entrevistador

Tabla: interview_comments.csv

Campo	Mnemónico asignado	Descripción
V1	llave	Llave de entrevista asignada por el sistema
V2	folio_sist	Folio asignado al registro por el sistema
V6	estatus	Estado de la entrevista al momento de la descarga
V8	fecha_com	Fecha en la que se registró el comentario
V9	hora_com	Hora en la que se registró el comentario
V10	autor_com	Figura operativa que realizó el comentario
V12	comentario	Comentario

Tabla: interview_diagnostics.csv

Campo	Mnemónico asignado	Descripción
V1	llave	Llave de entrevista asignada por el sistema
V2	folio_sist	Folio asignado al registro por el sistema
V4	entrev	Nominativo del entrevistador que levantó el cuestionario
V10	durac	Duración de la entrevista en minutos

Tabla: lista_residuos.csv

Campo	Mnemónico asignado	Descripción
V1	llave	Llave de entrevista asignada por el sistema
V2	folio_sist	Folio asignado al registro por el sistema
V3	tipo_resid	Código correspondiente al tipo de residuo declarado
V4	num_resid	Número de residuos declarados

Tabla: lista_tiporesiduos.csv

Campo	Mnemónico asignado	Descripción
V1	llave	Llave de entrevista asignada por el sistema
V2	folio_sist	Folio asignado al registro por el sistema
V3	tipo_resid	Código correspondiente al tipo de residuo declarado
V4	consec_residuo	Consecutivo del residuo declarado. Dado que el sistema inicia la enumeración de este consecutivo en 0, se suma 1 a cada registro para inicializar la cuenta en 1
V5	nombre_residuo	Descripción del residuo reportado
V6	cantidad_resid	Cantidad del residuo reportado
V7	unidad_resid	Unidad de medida del residuo reportado
V8	otra_unidad	Especifica otra unidad de medida del residuo reportado
V9	cant_r_no_comes	Cantidad del residuo reportado que es no comestible
V10	unid_r_no_comes	Unidad de medida del residuo reportado que es no comestible
V11	otra_ur_no_comes	Especifica otra unidad de medida del residuo reportado que es no comestible
V12	t_colecta	Periodo en el que reúne los residuos
V13	otro_t_colecta	Especifica otro periodo en el que reúne los residuos
V14	causa_desecho	Principal causa por la que se producen los residuos o desechos
V15	destino	Principal destino de los residuos
V16	cant_banco	Cantidad de residuos que se destinan a Banco de alimentos / comedor público
V17	unidad_banco	Unidad de medida de la cantidad de residuos que se destinan a Banco de alimentos / comedor público
V19	otro_destino	Especifica otro principal destino de los residuos
V20	autoriza_imag	Autoriza fotografiar imagen de residuos
V21	imagen_resid	Identifica nombre de imagen de residuos en la BD

- 14) Se filtran y dejan para integrar la base de datos de explotación, únicamente aquellos casos en los que se consiguió entrevista completa (se generaran o no residuos). En esta etapa se excluyen todos aquellos registros que se encuentran en una o más de las siguientes situaciones:

14.1 Establecimientos que cerraron o están abandonados, ya no se localizan en el sitio, cambiaron de giro a alguna clase de actividad no relacionada con alimentos, o es un giro distinto relacionado con alimentos, pero no se captaron.

14.2 Establecimientos en los que no pudo entrevistarse al informante adecuado, o no abrieron.

14.3 Negativas a la entrevista, o que refieren que debe consultarse a su corporativo.

14.4 Registros con código 1 pruebas de configuración del instrumento. 14.5 Registros rechazados por el supervisor

- 15) En cualquiera de los casos anteriores, el filtro a aplicar toma la siguiente expresión: **establecimientos1.acepta = 1 and establecimientos1.tipo = 2 and interview_comments.estatus <> @@RejectedBySupervisor**, es decir, todos aquellos establecimientos en los que se aceptó la entrevista, durante el operativo real de la prueba piloto, que fueron aceptados.
- 16) Una vez filtrados los registros válidos, se obtiene que la tabla *establecimientos1.csv* queda con un total de **120 registros, y 27 campos**.
- 17) A esta tabla, se le adicionan mediante *JOIN* los campos provenientes de la muestra: *nombre_establecimiento*, *codigo_scian*, *personas_ocupadas*, *municipio*, *tipo_unidad_economica*, *sector_scian*, *tipo_actividad*
- 18) La tabla resultante se guarda con el nombre *establecimientos2.csv*. En la misma se hacen algunos cambios dirigidos, según lo siguiente:

Llave del registro o condición lógica	Descripción del cambio realizado
53-27-23-59	tipo_unidad_economica = Fijo
53-27-23-59	sector_scian = Terciario
53-27-23-59	tipo_actividad = Restaurantes / servicio de preparación de alimentos
53-27-23-59	personas_ocupadas = 0 a 5 personas
53-27-23-59	municipio = Aguascalientes
66-30-95-73	tipo_unidad_economica = Fijo
75-88-19-43	tipo_unidad_economica = Fijo
75-88-19-43	sector_scian = Terciario
75-88-19-43	personas_ocupadas = 0 a 5 personas
75-88-19-43	municipio = Aguascalientes
58-57-15-91	tipo_unidad_economica = Fijo
77-06-67-30	tipo_unidad_economica = Fijo
98-38-33-96	tipo_unidad_economica = Fijo
id_denue >= 99999133	tipo_unidad_economica = Ambulante
id_denue >= 99999003	sector_scian = Terciario
id_denue >= 99999003	personas_ocupadas = 0 a 5 personas
id_denue >= 99999003	municipio = Aguascalientes
66-30-95-73	municipio = Aguascalientes
66-30-95-73	sector_scian = Terciario

- 19) Se obtienen las tablas restantes *interview_diagnostics2.csv*, *lista_residuos2.csv*, y *lista_tiporesiduos2.csv* depuradas con sólo registros de entrevistas válidas, y complementadas con variables adicionales de la muestra (*interview_comments* no se modifica). Se generan con las variables de la muestra, mediante un operador JOIN. Estas cuatro tablas conforman la **Base de Datos para explotación**.

20) La base para explotación queda conformada entonces de la siguiente forma:

Tabla	Número de registros	Número de campos
establecimientos2.csv	120	34
interview_diagnostics2.csv	119	11
lista_residuos2.csv	152	10
lista_tiporesiduos2.csv	191	31

De las 120 entrevistas logradas, 89 corresponden a establecimientos fijos o semifijos, y 31 a establecimientos ambulantes. La distribución por municipio fue la siguiente:

Municipio	Número de entrevistas
Aguascalientes	108
Calvillo	4
Jesús María	4
San Francisco de los Romo	2
Pabellón de Arteaga	2