

# ERAMO Prueba Piloto 2021

Bitácora de Procesamiento de la Base de Datos de Captura

Elaborada por: Raúl Mejía González

## Objetivo:

El objetivo de esta bitácora es llevar un registro preciso y puntual de cada de 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 las tablas (tratamientos dirigidos).

## 2. Procedimientos aplicados.

A continuación, se incluye el código R que se aplicó a cada una de las tablas para obtener la base de datos validada.

### Cuestionario Agropecuario

```
#### Tablas del cuestionario agropecuario:

##### Tratamiento de las bases de datos de salida, para #####
##          obtener bases de datos validadas.          ##

setwd("D:/inegi/eramo/prueba_piloto_2020/bases_de_datos/
      bd_eramo_pp2020_validada/agropecuario")

## Insumo: tablas csv de la BD antes de validación

##### Tabla eramo #####

eramo <- read.csv("eramo.csv")
dim(eramo)

# elimina campos innecesarios para el procesamiento:
```

```

eramo <- eramo[,-c(2,6:8,19,20,27,28,36:48,54,
                  55,64,65,69:85,91,92,94:105,107:110)]

dim(eramo)

# cambia nombres de campos:
names(eramo)[15] <- "registro_agr"
names(eramo)[16] <- "reg_agr_p1"
names(eramo)[17] <- "reg_agr_p2"
names(eramo)[18] <- "reg_agr_p3"
names(eramo)[19] <- "reg_agr_p4"
names(eramo)[20] <- "reg_agr_p5"
names(eramo)[21] <- "prod_agr_1"
names(eramo)[22] <- "prod_agr_2"
names(eramo)[23] <- "prod_agr_3"
names(eramo)[24] <- "prod_agr_4"
names(eramo)[25] <- "prod_agr_5"
names(eramo)[26] <- "prod_agr_6"
names(eramo)[27] <- "prod_agr_7"
names(eramo)[33] <- "obs_agr_1"
names(eramo)[34] <- "obs_agr_2"
names(eramo)[35] <- "registro_gan"
names(eramo)[36] <- "reg_gan_p1"
names(eramo)[37] <- "reg_gan_p2"
names(eramo)[38] <- "reg_gan_p3"
names(eramo)[39] <- "reg_gan_p4"
names(eramo)[40] <- "reg_gan_p5"
names(eramo)[41] <- "prod_gan_1"
names(eramo)[42] <- "prod_gan_2"
names(eramo)[43] <- "prod_gan_3"
names(eramo)[49] <- "obs_gan_1"
names(eramo)[50] <- "obs_ent"

names(eramo)

# Elimina registros sin información
eramo <- eramo[-which(eramo$interview__key %in% c("47-57-77-73",
                                                "27-62-48-05",
                                                "51-69-57-33")),]

# Elimina registros con negativa
eramo <- eramo[-which(eramo$ok_entrev == 3),]

# Coloca datos de georreferenciación registros:

eramo$coordenada__Longitude[which(eramo$interview__key == "06-17-29-37")] <-
-100-39/60
eramo$coordenada__Latitude[which(eramo$interview__key == "06-17-29-37")] <-
24+24/60

eramo$coordenada__Longitude[which(eramo$interview__key == "23-92-41-18")] <-
-100-54/60
eramo$coordenada__Latitude[which(eramo$interview__key == "23-92-41-18")] <-

```

```
##### Tabla interview_comments_erao: #####

comentarios <- read.csv("interview__comments_erao.csv")

# elimina campos innecesarios para el procesamiento:
comentarios <- comentarios[,-c(2,4,5,11)]

# Nombra variables
names(comentarios) <- c("interview__key","seccion","variable","num",
                        "fecha", "hora_ini", "cve_entrevistador", "comentario")

# Modifica textos de referencia a secciones

comentarios$seccion[which(comentarios$seccion == "lista_residuos")] <-
  "prod_agr"
comentarios$seccion[which(comentarios$seccion == "reg_pc_perdidas")] <-
  "perdidas_agr"
comentarios$seccion[which(comentarios$seccion == "reg_pp_perdidas")] <-
  "perdidas_gan"
comentarios$seccion[which(comentarios$seccion == "lista_crias")] <-
  "prod_gan"

##### Tabla interview_actions_erao: #####

acciones <- read.csv("interview__actions_erao.csv")

# elimina campos innecesarios para el procesamiento:
acciones <- acciones[,-c(2,5,7:9)]

# Nombra variables
names(acciones) <- c("interview__key","fecha", "hora_acc", "cve_entrevistador")

##### Tabla lista_residuos: #####

prod_agr <- read.csv("lista_residuos.csv")
#prod_agr <- read.csv("D:/inegi/erao/prueba_piloto_2020/bases_de_datos/
bd_erao_pp2020_validada/agropecuario/lista_residuos.csv")

# elimina campos innecesarios para el procesamiento:
prod_agr <- prod_agr[,-c(2,10,13,14,21,22,26,30)]
```

```

# Nombra variables

names(prod_agr) <- c("interview__key","num_prod","tipo_prod",
                    "cosecha_anual", "u_cosecha_anual",
                    "otra_unid_c_anual",
                    "consum_humano", "consum_animal",
                    "porc_humano", "porc_animal",
                    "mercad_nac_min", "mercad_nac_may", "mercad_extra",
                    "por_nac_min", "por_nac_may", "por_extra",
                    "cosecha_estimada",
                    "u_cosecha_estimada", "ha_cosecha_estimada",
                    "cosecha_final",
                    "u_cosecha_final", "ha_cosecha_final",
                    "pro_c_perdida1","pro_c_perdida2","pro_c_perdida3",
                    "pro_c_perdida4","pro_c_perdida5","pro_c_perdida6",
                    "otro_pro_c_perdida")

# Elimina registros sin información
prod_agr <- prod_agr[-which(prod_agr$interview__key %in% c("89-64-98-87",
"25-83-35-42","51-32-34-93")),]

##### Tabla reg_pc_perdidas: #####

perdidas_agr <- read.csv("reg_pc_perdidas.csv")
#perdidas_agr <- read.csv("D:/inegi/eramo/prueba_piloto_2020/bases_de_datos/
bd_eramo _pp2020_validada/agropecuario/reg_pc_perdidas.csv")

# elimina campos innecesarios para el procesamiento:
perdidas_agr <- perdidas_agr[,-2]

# Nombra variables

names(perdidas_agr) <- c("interview__key","num_prod",
                        "pro_c_perdida", "cant_per", "u_med_per",
                        "ha_proceso_perd", "otra_u_med_per",
                        "cant_mon_per", "causa", "destino")

# Elimina registros sin información
perdidas_agr <- perdidas_agr[-which(perdidas_agr$interview__key ==
"76-56-81-14"),]

##### Tabla lista_crias #####

prod_gan <- read.csv("lista_crias.csv")

```

```

# elimina campos innecesarios para el procesamiento:
prod_gan <- prod_gan[,-c(2,7,10:14,21,22,25,28)]

# Nombra variables

names(prod_gan) <- c("interview__key","num_prod","tipo_prod",
                    "prod_anual", "u_prod_anual",
                    "c_humano", "c_animal",
                    "c_mercad_nac_min",
                    "c_mercad_nac_may", "c_mercad_extranjero",
                    "c_por_nac_min",
                    "c_por_nac_may", "c_por_extranjero",
                    "prod_estimada", "u_prod_estimada",
                    "prod_final", "u_prod_final",
                    "pro_p_perdida1", "pro_p_perdida2",
                    "pro_p_perdida3", "pro_p_perdida4",
                    "pro_p_perdida5", "pro_p_perdida6",
                    "otro_pro_p_perdida")

##### Tabla reg_pp_perdidas #####

perdidas_gan <- read.csv("reg_pp_perdidas.csv")
#perdidas_gan <- read.csv("D:/inegi/eramo/prueba_piloto_2020/bases_de_datos/
bd_eramo_pp2020_validada/agropecuario/reg_pp_perdidas.csv")

# elimina campos innecesarios para el procesamiento:
perdidas_gan <- perdidas_gan[,-2]

# Nombra variables

names(perdidas_gan) <- c("interview__key", "num_prod", "pro_p_perdida",
                        "cant_p_per", "u_med_p",
                        "otra_u_med_p", "cant_mon_per_p", "causa_p",
                        "destino_p")

# Elimina registros sin información
perdidas_gan <- perdidas_gan[-which(perdidas_gan$interview__key %in%
c("06-17-29-37","09-39-85-34")
| (perdidas_gan$interview__key == "51-30-77-57" &
    perdidas_gan$num_prod %in% c(2,3))),]

write.csv(eramo,"agropecuario.csv")
write.csv(acciones,"acciones_entrevistas.csv")
write.csv(comentarios,"comentarios.csv")
write.csv(prod_agr,"productos_agricolas.csv")
write.csv(perdidas_agr,"perdidas_prod_agricolas.csv")
write.csv(prod_gan,"productos_pecuarios.csv")
write.csv(perdidas_gan,"perdidas_prod_pecuarios.csv")

```

## Cuestionario Establecimientos

```
##### Tratamiento de las bases de datos de salida, para #####
## obtener bases de datos validadas. ##

setwd("D:/inegi/eramo/prueba_piloto_2020/bases_de_datos/
      bd_eramo_pp2020_validada/establecimientos")

## Insumo: tablas csv de la BD antes de validación

##### Tabla desperdicio #####

establecimientos <- read.csv("desperdicio.csv")

# concatena campos de observaciones
establecimientos$obs <- paste(establecimientos$obs_ent_c,
                             establecimientos$V70_0, sep="")

# elimina campos innecesarios para el procesamiento:
establecimientos <- establecimientos[,-c(2,6:8,42:50,56:76)]

# corrige otra modalidad
establecimientos$otra_modalidad[which(establecimientos$interview__key ==
                                     "30-93-43-10")] <- "presencial, en papel"
establecimientos$otra_modalidad[which(establecimientos$interview__key ==
                                     "76-24-26-14")] <- "presencial, en papel"
establecimientos$otra_modalidad[which(establecimientos$interview__key ==
                                     "10-79-05-02")] <- "presencial, en papel"
establecimientos$otra_modalidad[which(establecimientos$interview__key ==
                                     "48-02-88-02")] <- "presencial, en papel"
establecimientos$otra_modalidad[which(establecimientos$interview__key ==
                                     "31-42-14-83")] <- "presencial, en papel"
establecimientos$otra_modalidad[which(establecimientos$interview__key ==
                                     "65-34-10-09")] <- "presencial, en papel"
establecimientos$otra_modalidad[which(establecimientos$interview__key ==
                                     "91-74-34-96")] <- "presencial, en papel"
establecimientos$otra_modalidad[which(establecimientos$interview__key ==
                                     "90-77-74-61")] <- "presencial, en papel"
establecimientos$otra_modalidad[which(establecimientos$interview__key ==
                                     "20-35-84-30")] <- "presencial, en papel"
establecimientos$otra_modalidad[which(establecimientos$interview__key ==
                                     "88-17-81-08")] <- "presencial, en papel"

# Elimina registros sin información (negativa o no dieron datos de productos)
establecimientos <- establecimientos[-which(establecimientos$interview__key %in%
                                             c("04-65-89-03",
                                                "28-16-90-16", "11-99-56-10", "41-15-60-61",
                                                "20-34-44-65", "47-22-19-75", "93-81-73-58",
```

```
"07-31-10-61","18-25-26-96","75-81-04-87"))],]
```

*# Coloca datos de georreferenciación registros:*

```
establecimientos$coordenada__Longitude[which(establecimientos$interview__key ==  
"77-01-74-26")] <- -102.8044149  
establecimientos$coordenada__Latitude[which(establecimientos$interview__key ==  
"77-01-74-26")] <- 25.3325315
```

```
establecimientos$coordenada__Longitude[which(establecimientos$interview__key ==  
"96-31-44-15")] <- -99.1278843  
establecimientos$coordenada__Latitude[which(establecimientos$interview__key ==  
"96-31-44-15")] <- 23.7198343
```

```
establecimientos$coordenada__Longitude[which(establecimientos$interview__key ==  
"74-51-28-39")] <- -99.1281397  
establecimientos$coordenada__Latitude[which(establecimientos$interview__key ==  
"74-51-28-39")] <- 23.7198294
```

```
establecimientos$coordenada__Longitude[which(establecimientos$interview__key ==  
"76-26-19-30")] <- -100.5335  
establecimientos$coordenada__Latitude[which(establecimientos$interview__key ==  
"76-26-19-30")] <- 28.7069167
```

```
establecimientos$coordenada__Longitude[which(establecimientos$interview__key ==  
"20-51-32-82")] <- -102.5734749  
establecimientos$coordenada__Latitude[which(establecimientos$interview__key ==  
"20-51-32-82")] <- 22.7746184
```

```
establecimientos$coordenada__Longitude[which(establecimientos$interview__key ==  
"26-55-87-94")] <- -102.5475399  
"26-55-87-94")] <- 22.7602031
```

```
establecimientos$coordenada__Longitude[which(establecimientos$interview__key ==  
"12-33-17-05")] <- -99.1707222  
establecimientos$coordenada__Latitude[which(establecimientos$interview__key ==  
"12-33-17-05")] <- 19.688667
```

```
establecimientos$coordenada__Longitude[which(establecimientos$interview__key ==  
"02-22-73-77")] <- -99.6866389  
establecimientos$coordenada__Latitude[which(establecimientos$interview__key ==  
"02-22-73-77")] <- 19.689194
```

```
establecimientos$coordenada__Longitude[which(establecimientos$interview__key ==  
"09-55-11-13")] <- -99.166361  
establecimientos$coordenada__Latitude[which(establecimientos$interview__key ==  
"09-55-11-13")] <- 19.6891667
```

```
establecimientos$coordenada__Longitude[which(establecimientos$interview__key ==  
"20-27-37-35")] <- -99.16575
```

```

establecimientos$coordenada__Latitude[which(establecimientos$interview__key ==
"20-27-37-35")] <- 19.688972

establecimientos$coordenada__Longitude[which(establecimientos$interview__key ==
"97-73-87-30")] <- -99.1659444
establecimientos$coordenada__Latitude[which(establecimientos$interview__key ==
"97-73-87-30")] <- 19.6904167

establecimientos$coordenada__Longitude[which(establecimientos$interview__key ==
"53-73-04-92")] <- -99.1658889
establecimientos$coordenada__Latitude[which(establecimientos$interview__key ==
"53-73-04-92")] <- 19.6906667

establecimientos$coordenada__Longitude[which(establecimientos$interview__key ==
"47-84-82-42")] <- -99.1497096
establecimientos$coordenada__Latitude[which(establecimientos$interview__key ==
"47-84-82-42")] <- 19.4787908

establecimientos$coordenada__Longitude[which(establecimientos$interview__key ==
"31-28-80-45")] <- -99.1538724
establecimientos$coordenada__Latitude[which(establecimientos$interview__key ==
"31-28-80-45")] <- 19.4763635

establecimientos$coordenada__Longitude[which(establecimientos$interview__key ==
"44-35-84-50")] <- -99.09650
establecimientos$coordenada__Latitude[which(establecimientos$interview__key ==
"44-35-84-50")] <- 19.65439

establecimientos$coordenada__Longitude[which(establecimientos$interview__key ==
"91-51-53-80")] <- -99.09657
establecimientos$coordenada__Latitude[which(establecimientos$interview__key ==
"91-51-53-80")] <- 19.64841

establecimientos$coordenada__Longitude[which(establecimientos$interview__key ==
"49-55-81-49")] <- -99.09360
establecimientos$coordenada__Latitude[which(establecimientos$interview__key ==
"49-55-81-49")] <- 19.65283

establecimientos$coordenada__Longitude[which(establecimientos$interview__key ==
"83-07-51-12")] <- -99.11048
establecimientos$coordenada__Latitude[which(establecimientos$interview__key ==
"83-07-51-12")] <- 19.65508

establecimientos$coordenada__Longitude[which(establecimientos$interview__key ==
"25-40-36-43")] <- -99.09838
establecimientos$coordenada__Latitude[which(establecimientos$interview__key ==
"25-40-36-43")] <- 19.66160

```

```
##### Tabla interview_comments_desperdicios: #####
```



```

comentarios <- read.csv("interview__comments_desperdicios.csv")

# elimina campos innecesarios para el procesamiento:
comentarios <- comentarios[,-c(2,4,10)]

# Nombra variables
names(comentarios) <- c("interview__key","seccion","variable","num",
                        "fecha", "hora_ini", "cve_entrevistador", "comentario")

##### Tabla interview_actions_desperdicios: #####

acciones <- read.csv("interview__actions_desperdicios.csv")

# elimina campos innecesarios para el procesamiento:
acciones <- acciones[,-c(2,5,7:9)]

# Nombra variables
names(acciones) <- c("interview__key","fecha", "hora_acc", "cve_entrevistador")

##### Tabla lista_residuos: #####

prod_estab <- read.csv("lista_productos.csv")

# elimina campos innecesarios para el procesamiento:
prod_estab <- prod_estab[,-c(2,9)]

# asigna dato recuperado de pérdida
prod_estab$perdida_prod[which(establecimientos$interview__key == "20-71-86-36")]
<- 2

# elimina registros sin datos numericos de manejo y de pérdidas
prod_estab <- prod_estab[-which(prod_estab$interview__key %in% c("02-85-64-84",
                                                                "54-20-67-43",
                                                                "01-43-03-64",
                                                                "56-93-50-65",
                                                                "09-43-30-81",
                                                                "16-17-40-85",
                                                                "26-55-87-94",
                                                                "10-79-05-02",
                                                                "97-40-86-36",
                                                                "74-62-60-95",
                                                                "88-17-81-08",
                                                                "90-77-74-61",
                                                                "18-58-26-99",

```

```
"20-51-32-82",  
"18-50-52-79"))],]
```

```
# Renombra variables
```

```
names(prod_estab)[2] <- "num_prod"  
names(prod_estab)[10] <- "din_perd"  
names(prod_estab)[12] <- "comestible"  
names(prod_estab)[13] <- "u_comestible"  
names(prod_estab)[14] <- "otra_u_comestible"  
  
write.csv(establecimientos,"establecimientos.csv")  
write.csv(acciones,"acciones_entrevistas.csv")  
write.csv(comentarios,"comentarios.csv")  
write.csv(prod_estab,"productos_establecimientos.csv")
```