



INSTITUTO NACIONAL DE ESTADÍSTICA



ENCUESTA DE POBLACIÓN ACTIVA

Curso nuevo ingreso

24 junio 2016

Gestión y tratamiento centralizado de los datos



Tratamiento de la información

☐ En la Delegación Provincial o Centros CATI

- Realización de la entrevista
- Revisión de la entrevista

☐ En los Servicios Centrales

- Depuración manual
- Depuración automática
- Generación de variables derivadas y cálculo de los factores de elevación
- Obtención de los productos de difusión

Proceso seguido por los cuestionarios (semanal)

- ☐ Realización de entrevistas en CATI / CAPI
- ☐ Codificación, revisión y depuración manual en la Delegación Provincial o centro de origen (CATI / CAPI)
- ☐ Descarga para trasvase a Servicios Centrales
- ☐ Incorporación de viviendas de semanas anteriores (hasta 6 semanas)
- ☐ Proceso en Servicios Centrales

Fases del proceso en Servicios Centrales (mensual/trimestral)

- ❑ Detección automática-corrección manual (semanas 1 a 4, 5 a 8 y 9 a 13)
- ❑ Proceso trimestral con incorporación de atrasados que se han recibido después del correspondiente proceso mensual
- ❑ Publicación de los resultados

Objetivos del tratamiento informático

- ☐ Depurar los datos
 - Llegar al *valor verdadero* (depuración manual)
 - Dejar los datos *coherentes* (depuración automática)
- ☐ Incorporar todos los elementos necesarios para la explotación del fichero

Detección automática. Depuración manual

- ☐ Comprobación de la estructura de grabación
- ☐ Detección viviendas duplicadas y errores de cobertura. Contraste con diccionario geográfico
- ☐ Errores en la estructura del hogar y en las variables de control

EI DIA

- ❑ Aplicación informática desarrollada por el INE para la depuración de variables cualitativas
- ❑ Basado en la metodología de Fellegi y Holt con modificaciones para manejar errores sistemáticos
- ❑ Permite sólo detección o detección más imputación

Ventajas del DIA

- ❑ Facilidad de manejo: sólo hay que definir las inconsistencias (edits o rids)
- ❑ Paquete integrado y probado
- ❑ Flexible: permite modificar fácilmente las reglas
- ❑ Proporciona gran cantidad de información complementaria: imputaciones, distribuciones, porcentaje de errores, etc.

Inconvenientes del DIA

- ❑ No admite reglas de inconsistencia entre variables de registros diferentes
- ❑ Problemas para el tratamiento de características que pueden tomar muchos valores diferentes
- ❑ Consume mucho tiempo de proceso y memoria del ordenador

Ejemplo de edit de flujo

DUCON1(6)NENT(6)MOTEMP(b)

siendo DUCON1 la variable que se refiere a los tipos de contrato

MOTEMP, variable que pregunta acerca del motivo por el que se tiene un contrato temporal

NENT la variable que indica el número de entrevista

DUCON1=1 => Contrato indefinido

DUCON1=6 => Contrato temporal

Ejemplo de edit de contenido

EMPANT(6)NBUSCA(2,7)

siendo EMPANT la variable que se refiere a la realización de algún trabajo anteriormente.

NBUSCA, variable de razones por las que no busca empleo

EMPANT=6 => No ha trabajado anteriormente

NBUSCA=2,7 => No busca empleo por:

- Afectado por regulación
- Está jubilado



El DIA en la EPA

- ☐ DIA para el cuestionario de menores
- ☐ DIA para el cuestionario de personas de 16 o más años

El DIA de los mayores

- ❑ Dividido en cinco fases
- ❑ Se han definido unas *variables auxiliares* que resumen apartados completos del cuestionario
- ❑ Cálculo de duraciones en lugar de fechas

Fases posteriores a la depuración (1/3)

- ❑ Copia de registros del trimestre anterior
 - Paliar la falta de respuesta
 - Debe existir información en t-1
 - La información de t-1 no puede ser copia de t-2

Fases posteriores a la depuración (2/3)

- ❑ Validación del fichero de secciones a repetir del trimestre de proceso
- ❑ Validación del fichero de poblaciones del trimestre a publicar

Fases posteriores a la depuración (3/3)

- ❑ Asignación de los *códigos de estrato* y de los *factores de repetición y de elevación* al fichero de unión de *contestados y copiados*
- ❑ Asignación de variables derivadas
- ❑ Creación del fichero final de microdatos trimestral

Disponibilidad de las estimaciones (EPA trimestral)

- ☐ A comienzo del año: calendario de disponibilidad
- ☐ Trimestre 1: 4ª semana de abril
- ☐ Trimestre 2: 4ª semana de julio
- ☐ Trimestre 3: 4ª semana de octubre
- ☐ Trimestre 4: 4ª semana de enero del año siguiente

Datos EPA en la web

- ❑ [Página web del INE](#)

Muchas gracias por vuestra atención

M. Flor Gil Alonso