



VII ENCUESTA DE PRESUPUESTOS FAMILIARES

**MANUAL DE USUARIO DE BASE DE DATOS
(STATA)**

INSTITUTO NACIONAL DE ESTADÍSTICAS

**Septiembre / 2013
N° 01**

VII ENCUESTA DE PRESUPUESTOS FAMILIARES SUBDIRECCIÓN TÉCNICA

Manual de Usuario (Stata)
Instituto Nacional de Estadísticas.
Septiembre / 2013.
Nº 01

Jefe de Proyecto: Francisco Javier Bilbao Quiroga
Encargada Equipo Técnico: Rocío De Los Ángeles Miranda Rocco
Analistas Equipo Técnico: Luis Roberto Figueroa Mendizábal
Leonardo Jaime González Allendes
Paula Eugenia Lara Martínez
Maribel Valesca Maquieira Veloso
Patricia Alejandra Mauna Madrid
Guillermo Antonio Pazols Melgarejo
Beatriz Salinas Quiroga

ÍNDICE

1.	PRESENTACIÓN	3
2.	ACCESO Y SOLICITUD DE LOS DATOS	3
3.	ESTRUCTURA DE LOS DATOS	4
	3.1. Base de Personas (BASE_PERSONAS_VIEPF)	4
	3.2. Base de Gasto (BASE_GASTOS_VIEPF)	5
	3.3. Base Clasificador de Productos (CCIF_VIEPF)	6
4.	ANÁLISIS Y TABULADOS	8
	4.1. Identificadores y Tablas	8
	4.2. Consolidación de las bases	20
	4.3. Quintilización por ingreso	21

1. PRESENTACIÓN

El presente manual está destinado a los usuarios de las bases de datos de la VII Encuesta de Presupuestos Familiares (VII EPF) aplicada entre noviembre de 2011 y octubre de 2012.

En este manual es posible encontrar orientaciones para el manejo y análisis de las bases de datos, así como los principales tabulados para los módulos de caracterización socio-demográfica y ocupacional del hogar, ingresos y tipo de gastos.

Para el trabajo y tratamiento de la información, se utiliza el programa estadístico-econométrico Stata.

La definición de conceptos, su tratamiento estadístico y metodología de captura, se encuentran detallados en los documentos publicados por la encuesta y la Metadata disponible en el portal del Instituto Nacional de Estadística (INE)¹.

2. ACCESO Y SOLICITUD DE LOS DATOS

Las bases de datos de la VII EPF se encuentran publicadas en el portal del Instituto Nacional de Estadística (INE)² junto a los otros documentos elaborados. Para acceder a los datos publicados por la encuesta, los usuarios deben completar un registro de solicitud en el portal para descargar las bases en formato Stata o SPSS. También pueden solicitar la información directamente en las oficinas de la institución.

Aquellos usuarios de la encuesta que requieran información detallada o tengan consultas sobre el uso de la base de datos, pueden realizar las solicitudes y preguntas a través de la Oficina de Informaciones, Reclamos y Sugerencias (OIRS) de la institución. La información adicional solicitada por los usuarios, puede contemplar diversos módulos de la encuesta y niveles de desagregación. Sin embargo, la entrega de esta información se encuentra sujeta a los siguientes criterios:

¹ Se recomienda a los usuarios acompañar el análisis de ambas bases de datos, con la metodología de la encuesta, el manual de trabajo de campo y la Metadata o diccionario de variables.

² www.ine.cl/epf

- Disponibilidad de la información solicitada.
- Confidencialidad según la Ley del Secreto Estadístico 17.374.

La responsabilidad en el uso de la información con niveles de desagregación que no sean estadísticamente significativos, será del investigador y no de la institución (INE).

Los datos de ingreso y gasto de la encuesta tienen representatividad estadística sólo a nivel del **área de estimación de la muestra**³ en términos geográficos y con temporalidad anual.

3. ESTRUCTURA DE LOS DATOS

La información de la encuesta se encuentra publicada en tres bases de datos. La BASE_PERSONAS_VIIEPF contiene información sobre las personas y los hogares encuestados, la BASE_GASTOS_VIIEPF contiene información desagregada del gasto, a nivel de clasificación de productos CCIF, realizado por los hogares, mientras que la base CCIF_VIIEPF contiene el clasificador utilizado por la encuesta⁴.

3.1. Base de Personas (BASE_PERSONAS_VIIEPF)

La información de la base de personas se encuentra a nivel de hogares y personas, su estructura consiste en 61 variables y 35.669 observaciones. La información corresponde a las características sociodemográficas de los miembros del hogar y su situación ocupacional, la tenencia de la vivienda primaria y secundaria, el gasto total del hogar, los ingresos del hogar y sus fuentes.

³ La representatividad geográfica de la muestra comprende el total de capitales regionales, Gran Santiago y a nivel de Resto de Capitales Regionales. Para una explicación más detallada, revisar la metodología de la encuesta disponible en el sitio web del INE.

⁴ La VII Encuesta de Presupuestos Familiares para codificar sus productos utiliza la Clasificación del Consumo Individual por Finalidades (CCIF) que corresponde a una categorización funcional del sistema de cuentas nacionales y se encuentra estandarizada en sus niveles más agregados de clasificación por naciones unidas. Para una explicación más detallada, revisar la metodología de la encuesta.

3.1.1. La Información sociodemográfica del hogar se encuentra desglosada en 18 variables que describen algunas características de los miembros del hogar y la condición de tenencia de la vivienda (principal y secundaria).

3.1.2. La información sobre la actividad económica, salud y previsión social de los miembros del hogar, se encuentra contenida en 9 variables que describen principalmente la población ocupada, el tipo de ocupación y la posesión de sistema de salud Y/o previsión.

3.1.3. La información sobre los ingresos del hogar comprende 32 variables que describen el ingreso total, así como los ingresos según el tipo de fuente. Por otra parte, los ingresos del hogar y por ocupación principal corresponden al cálculo del ingreso disponible y bruto. El monto de los ingresos presentados considera tanto los ingresos observados, como los ingresos imputados.

3.1.4. La información sobre el gasto total del hogar, comprende 2 variables que corresponden al gasto para cada hogar encuestado, excluyendo e incluyendo el arriendo imputado.

3.2. Base de Gasto (BASE_GASTOS_VIIEPF)

Los datos presentados en dicha base, contemplan la información los gastos realizados por el hogar a nivel de producto, de acuerdo a la Clasificación de Consumo Individual por Finalidades (CCIF). Dicha base cuenta con 11 variables y 641.938 observaciones.

3.2.1. La identificación del hogar, el área geográfica y el factor de expansión comprenden 3 variables específicas.

3.2.2. La información del gasto del hogar desagregado por CCIF, comprende 8 variables que especifican las subdivisiones del clasificador y el monto del gasto por producto.

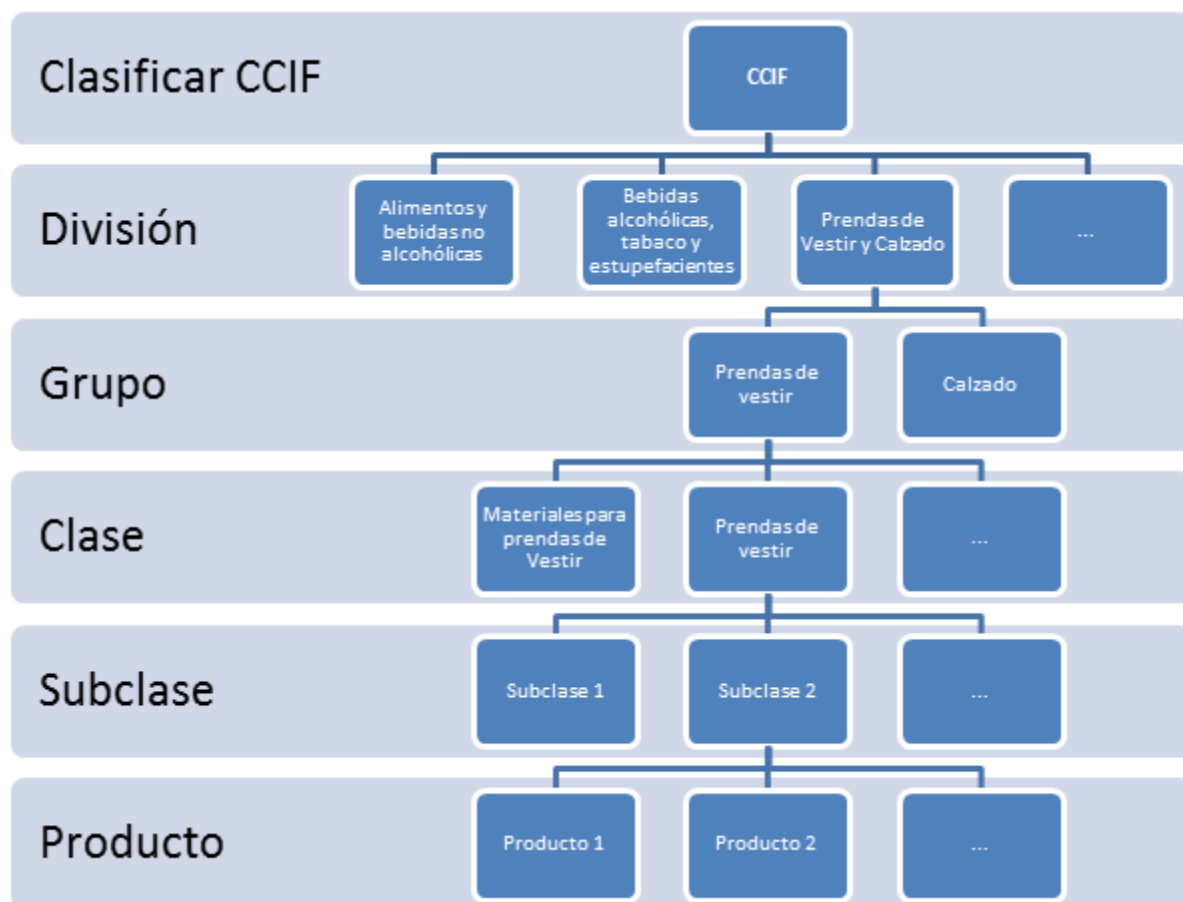
3.3. Base Clasificador de Productos (CCIF_VIIEPF)

Esta base de datos contiene información sobre la clasificación de bienes y servicios utilizada en la VII EPF. La base de datos contiene 5 niveles de desagregación en la base de datos. Cada producto se identifica por un código único.

La CCIF consta de 12 divisiones y la información se encuentra desagregada en 5 niveles. La VII EPF en versión CCIF utiliza la misma clasificación que el IPC hasta su tercer nivel de desagregación (División, Grupo, Clase), ya que estas son normadas a nivel internacional, sin embargo, los últimos dos niveles de clasificación, no coinciden en el detalle con la clasificación utilizada por IPC, ya que la EPF abarca una mayor variedad de bienes y servicios.

Según mayor a menor agregación, los bienes y servicios se ordenan en 12 divisiones, las que a su vez se subdividen en grupos y estos se vuelven a subdividir en clases. Estos tres niveles de división del clasificador son fijados por las Naciones Unidas y permiten la comparabilidad internacional de forma unificada, sin embargo, las clases son nuevamente subdivididas en subclases y estas a su vez se dividen en productos.

El diagrama muestra un ejemplo del funcionamiento del clasificador CCIF con sus diferentes niveles de clasificación de los gastos.



Las siguientes variables muestran las desagregaciones:

D: nivel de divisiones (12 divisiones)

G: nivel de grupo (59 grupos)

C: nivel de clases (126 clases)

SC: nivel de subclases (274 subclases)

P: nivel de productos (1099 productos)

CCIF: es la unión de todos los anteriores niveles y el dígito se compone de la siguiente forma "D.G.C.SC.P"

4. ANÁLISIS Y TABULADOS

Considerando la estructura de ambas bases de datos y la información que contienen, se detallan algunos tabulados principales que los usuarios pueden replicar. Es importante mencionar que no se detallan todos los tabulados posibles a obtener a partir de los datos de la encuesta y que las programaciones detalladas no son la única forma de obtener los datos, se busca guiar a los analistas en el manejo y tratamiento de la información con el fin de reducir las complejidades de la base de datos⁵.

4.1. Identificadores y Tablas

- Expandir la población

La expansión de los datos de la encuesta a la población, se calcula utilizando la variable FE.

```
Stata  
table ZONA [pweight=FE], row f(%15.0f)
```

- Identificación de hogares

La identificación del número de hogares expandido en la BASE_PERSONAS_VIIEPF se calcula a partir de la variable JHOGAR, la cual indica el Jefe(a) de Hogar. Para realizar cualquier otro tipo de cálculo que considere sólo los hogares, se puede utilizar como filtro.

```
Stata  
table JHOGAR if JHOGAR!=0 [pweight=FE] f(%15.0f)
```

⁵ A partir del lenguaje de programación, los tabulados pueden ser obtenidos de diversas formas, el manual sólo entrega una programación de carácter referencial para el tratamiento de la información.

En la BASE_GASTOS_VIIEPF el cálculo del número de hogares se debe realizar a partir de la variable FOLIO, se construye una variable que contabilice el número de observaciones por hogar y luego se asigna el valor mínimo (dejando sólo una observación por hogar).

Stata

```
bysort FOLIO: gen contar=_n  
bysort FOLIO: gen hogar=1 if contar==1  
tab hogar //muestra el número total de hogares en la encuesta
```

- Número personas del hogar

El número de personas del hogar es una variable que se encuentra contenida en la base de personas, pero se puede calcular creando una variable auxiliar que contabilice las personas del hogar y luego calcular el valor máximo de la variable auxiliar. Finalmente se calcula el número de personas considerando el total de hogares expandidos.

Stata

```
bysort FOLIO (PERSONA): gen contarp=_n  
bysort FOLIO : egen numerop=max(contarp)  
table numerop if JHOGAR==1 [pweight=FE], row f(%15.0f)
```

- Tenencia de vivienda principal y secundaria en los hogares

La información sobre la tenencia de la vivienda primaria y secundaria se replica para todos los miembros del hogar, por ello, para calcular el tipo de tenencia, valor de arriendo imputado y la existencia de tenencia secundaria, se requiere utilizar como filtro la variable JHOGAR.

Stata

```
table TVP if JHOGAR==1 [pweight=FE], row f(%15.0f)  
table VS if JHOGAR==1 [pweight=FE], row f(%15.0f)  
  
table TVP if JHOGAR==1 & AI_VP!=. [pweight=FE], cont(mean AI_VP) row  
f(%15.0f)  
table VS if JHOGAR==1 & AI_VS2!=. [pweight=FE], cont(mean AI_VS2) row  
f(%15.0f)
```

- Población de ocupados

La población de ocupados cuyos ingresos el mes anterior corresponden a la misma actividad que realiza en la semana de referencia, se puede calcular utilizando las variables CAEG y AEPRL como filtros.

Stata

```
table SEXO if CAEG==1 & AEPRL==1 [pweight=FE], row f(%15.0f)
```

- Ingresos medios por ocupación principal

Siguiendo el mismo proceso de cálculo de la población de ocupados cuyos ingresos el mes anterior corresponden a la misma actividad que realiza en la semana de referencia, se utilizan las variables CAEG y AEPRL como filtros. Adicionalmente se construye una variable que agrupa los ingresos disponibles por actividad principal para obtener la media.

Stata

```
replace INGDA_HD=0 if INGDA_HD ==.  
replace INGDH_HD=0 if INGDH_HD ==.  
replace INGDCP_HD=0 if INGDCP_HD ==.  
replace INGDPI_HD=0 if INGDPI_HD ==.  
gen double ingreso= INGDA_HD + INGDH_HD + INGDCP_HD + INGDPI_HD  
  
table SEXO if CAEG==1 & AEPRL==1 [pweight=FE], cont(mean ingreso) row  
f(%15.0f)
```

- Ingreso total por hogar sin arriendo imputado (Disponible y Bruto)

Para el cálculo de los ingresos disponibles o brutos del hogar se debe, en primer lugar, convertir todas las fuentes de ingreso a valores iguales o mayores de cero para poderlos sumar. Luego, de calcular el total por cada miembro del hogar, se realiza la agregación de los ingresos totales para el hogar. Finalmente aquellos hogares con ID_MISSING⁶ con valor 1 deben ser reemplazados con ingreso missing, debido a la subdeclaración de ingresos.

⁶ La variable ID_MISSING (=1) identifica a los hogares con ingresos inferiores a cero se deben a la subdeclaración de los ingresos, y que además declaran transferencias emitidas a otros hogares, o a Instituciones sin fines de lucro; así como también a los hogares que no declaran fuente de ingresos y no son sujetos de imputación.

Stata

```

use "BASE_PERSONAS_VIIEPF.dta", clear
//INGDHOG_HD "Ingreso disponible total por hogar (sin arriendo
imputado)"
foreach v in INGDA_HD INGDH_HD INGDCP_HD INGDPI_HD INGJ_HD INGOTA
INGOTI INGP INGF INGT_NETA {
replace `v'=0 if `v'==.
}
gen double INGDHOG_persona= INGDA_HD + INGDH_HD + INGDCP_HD +
INGDPI_HD + INGJ_HD*(0.93) + INGOTA + INGOTI + INGP + INGF +
INGT_NETA
bys FOLIO: egen double INGDHOG_HD =total(INGDHOG_persona)
replace INGDHOG_HD=. if ID_MISSING==1

//INGDHOG_HD_AI "Ingreso bruto total por hogar (sin arriendo
imputado)"
foreach v in INGA_HD INGH_HD INGCP_HD INGPI_HD INGJ_HD INGOTA
INGOTI INGP INGF INGT_NETA {
replace `v'=0 if `v'==.
}
gen double INGHOG_persona= INGA_HD + INGH_HD + INGCP_HD + INGPI_HD
+ INGJ_HD+ INGOTA + INGOTI + INGP + INGF + INGT_NETA
bys FOLIO: egen double INGHOG_HD =total(INGHOG_persona)
replace INGHOG_HD=. if ID_MISSING==1

```

- **Ingreso total por hogar con arriendo imputado (Disponible y Bruto)**

El procedimiento es similar al anterior, sin embargo, las partidas de arriendo imputado en la base se repiten para todos los miembros del hogar, por lo que antes de sumar se deben convertir los valores de los arriendos imputados de la vivienda principal y secundaria a cero para los miembros del hogar menos para el jefe de hogar.

Stata

```

//INGDHOG_HD_AI "Ingreso disponible total por hogar (con arriendo
imputado)"
foreach v in INGDA_HD INGDH_HD INGDCP_HD INGDPI_HD INGJ_HD INGOTA
INGOTI INGP INGF INGT_NETA {
replace `v'=0 if `v'==. | ID_MISSING==1
}
bys FOLIO (PERSONA): replace AI_VP=0 if PERSONA!=1 | AI_VP==.
bys FOLIO (PERSONA): replace AI_VS2=0 if PERSONA!=1 | AI_VS2==.

```

```

gen double INGDHOG_AI_persona= INGDA_HD + INGDH_HD + INGDGP_HD +
INGDPI_HD + INGJ_HD*(0.93) + INGOTA + INGOTI + INGP + INGF +
INGT_NETA+AI_VP+AI_VS2
bys FOLIO: egen double INGDHOG_HD_AI =total(INGDHOG_AI_persona)

//INGDHOG_HD_AI "Ingreso bruto total por hogar (con arriendo
imputado)"
foreach v in INGA_HD INGH_HD INGCP_HD INGPI_HD INGJ_HD INGOTA
INGOTI INGP INGF INGT_NETA {
replace `v'=0 if `v'==. | ID_MISSING==1
}
bys FOLIO (PERSONA): replace AI_VP=0 if PERSONA!=1 | AI_VP==.
bys FOLIO (PERSONA): replace AI_VS2=0 if PERSONA!=1 | AI_VS2==.
gen double INGHOG_persona= INGA_HD + INGH_HD + INGCP_HD + INGPI_HD
+ INGJ_HD+ INGOTA + INGOTI + INGP + INGF + INGT_NETA+AI_VP+AI_VS2
bys FOLIO: egen double INGHOG_HD_AI=total(INGHOG_persona)

```

- **Ingreso y Gasto per cápita (con y sin arriendo imputado)**

Las variables de Ingreso disponible total por hogar (con y sin arriendo imputado) y Gasto total por hogar (con y sin arriendo imputado) per cápita se calculan dividiendo las variables totales por hogar entre el número total de personas del hogar (NPERSONA).

```

Stata
gen double INGHOG_HD_PC=INGHOG_HD/NPERSONA
gen double INGHOG_HD_AI_PC= INGHOG_HD_AI /NPERSONA

gen double GASTOT_FNR_PC=GASTOT_FNR/ NPERSONA
gen double GASTOT_FNR_AI_PC=GASTOT_FNR_AI / NPERSONA

```

- Tabulados de Gasto e Ingreso por Características del Jefe de Hogar

Los siguientes tabulados muestran resultados sobre el gasto y el ingreso promedio por hogar para el total nacional según características del jefe de hogar.

Stata

```
use "BASE_PERSONAS_VIIEPF.dta", clear
//Según Sexo
tabstat GASTOT_FNR [w=FE] if PARENTESCO==1, stat(mean) f(%15.4f) miss
by(SEXO) //GASTO PROMEDIO POR HOGAR PARA EL TOTAL DE CAPITALS
REGIONALES POR HOGAR Y POR SEXOS
tabstat INGDHOG_HD [w=FE] if PARENTESCO==1, stat(mean) f(%15.4f) miss
by(SEXO) //INGRESO PROMEDIO POR HOGAR PARA EL TOTAL DE CAPITALS
REGIONALES POR HOGAR Y POR SEXOS
```

A partir de la base de datos de persona, se pueden obtener diferentes tabulados sobre las distintas especificaciones de las variables de interés según las características del jefe de hogar reemplazando la variable a tabular. Si se reemplaza en el espacio que utilizan las variables GASTOT_FNR ó INGHOG_HD por otras especificaciones de las mismas, se pueden obtener estadísticas del gasto/ingreso promedio mensual según las variables del jefe de hogar.

A continuación se agregan posibles combinaciones de estimaciones de gasto e ingreso a obtener según el sexo del jefe de hogar. Para utilizar esta programación, se debe ejecutar el comando que crea el gasto y el ingreso per cápita.

Stata

```
use "BASE_PERSONAS_VIIEPF.dta", clear
//Según Sexo
foreach x in GASTOT_FNR GASTOT_FNR_AI INGDHOG_HD INGDHOG_HD_AI
INGHOG_HD_PC INGHOG_HD_PC GASTOT_FNR_PC GASTOT_FNR_AI_PC {
tabstat `x' [w=FE] if PARENTESCO==1, stat(mean) f(%15.4f) miss by(SEXO)
}
```

Para tabular los gastos y los ingresos según tramos etarios, la variable tramos etarios debe ser construida. A continuación se detalla la construcción de la variable tramos etarios para tramos cada 5 años, siendo el último tramo de 65 años o más.

Stata

```
use "BASE_PERSONAS_VIIEPF.dta", clear
//Según Tramos Etarios
//Generar los Tramos Etarios (Cada 5 años)
gen TR_ED=1 if EDAD>=0 & EDAD<=4
replace TR_ED=2 if EDAD>=5 & EDAD<=9
replace TR_ED=3 if EDAD>=10 & EDAD<=14
replace TR_ED=4 if EDAD>=15 & EDAD<=19
replace TR_ED=5 if EDAD>=20 & EDAD<=24
replace TR_ED=6 if EDAD>=25 & EDAD<=29
replace TR_ED=7 if EDAD>=30 & EDAD<=34
replace TR_ED=8 if EDAD>=35 & EDAD<=39
replace TR_ED=9 if EDAD>=40 & EDAD<=44
replace TR_ED=10 if EDAD>=45 & EDAD<=49
replace TR_ED=11 if EDAD>=50 & EDAD<=54
replace TR_ED=12 if EDAD>=55 & EDAD<=59
replace TR_ED=13 if EDAD>=60 & EDAD<=64
replace TR_ED=14 if EDAD>=65
replace TR_ED=15 if (EDAD== -88 | EDAD== -99)

label var TR_ED "Tramos de Edad"
label define TR_ED 1 "0-4" 2 "5-9" 3 "10-14" 4 "15-19" 5 "20-24" 6 "25-29" 7
"30-34" 8 "35-39" 9 "40-44" 10 "45-49" 11 "50-54" 12 "55-59" /*
*/ 13 "60-64" 14 "65 o más" 15 "NO SABE/NO RESPONDE", modify
label values TR_ED TR_ED

//Tabular el gasto/ingreso promedio por hogar para el total capitales regionales
por tramos etarios del jefe de hogar
foreach x in GASTOT_FNR GASTOT_FNR_AI INGDHOG_HD INGDHOG_HD_AI {
  tabstat `x' [w=FE] if PARENTESCO==1, stat(mean) f(%15.4f) miss by(TR_ED)
}

//Tabular el gasto/ingreso promedio por hogar para el total capitales regionales
por tramos etarios del jefe de hogar y por sexo
foreach x in GASTOT_FNR GASTOT_FNR_AI INGDHOG_HD INGDHOG_HD_AI {
  tabstat `x' [w=FE] if PARENTESCO==1, stat(mean) f(%15.4f) miss by(TR_ED)
}
```

Para tabular el gasto/ingreso promedio del hogar para el total de capitales regionales según el nivel educativo del jefe de hogar, se debe tabular utilizando la variable EDUNIVEL.

Stata

```
use "BASE_PERSONAS_VIIEPF.dta", clear
//Según Sexo
foreach x in GASTOT_FNR GASTOT_FNR_AI INGDHOG_HD INGDHOG_HD_AI {
  bysort SEXO: tabstat `x' [w=FE] if PARENTESCO==1, stat(mean) f(%15.4f)
  miss by(EDUNIVEL)
}
```

Para conseguir otros tabulados según otras características del jefe de hogar, basta reemplazar distintas variables en la variable en negrita (**EDUNIVEL**). Por ejemplo es posible realizar tabulados según la CISE del jefe de hogar (se reemplaza la variable por **AECISE**), según tenencia de la vivienda principal (reemplazando por **TVP**) entre otros.

- **Tabulados de Gasto e Ingreso por Características del Jefe de Hogar para Gran Santiago y para el Resto de Regiones**

Para obtener los cuadros descritos con anterioridad a nivel de GRAN SANTIAGO o a nivel del RESTO REGIONES, se debe incorporar la ZONA a las variables a tabular.

A continuación se especifica el cálculo del gasto promedio por hogar incorporando la opción de ZONA (Gran Santiago y Resto de Regiones).

Stata

```
use "BASE_PERSONAS_VIIEPF.dta", clear
//Según Sexo
bysort ZONA: tabstat GASTOT_FNR [w=FE] if PARENTESCO==1, stat(mean)
f(%15.4f) miss by(SEXO)
```

Siguiendo la misma instrucción de los tabulados precedentes, se pueden obtener datos del gasto incluyendo arriendo imputado o de los ingresos, basta reemplazar la variable GASTO_FNR por la variable que se desee obtener.

- Tipo de gasto por hogar (CCIF)

Para calcular el gasto en sus distintas desagregaciones (y no sólo el gasto a nivel del hogar), se debe trabajar con la base de gastos.

A continuación se detalla los pasos que deben seguirse para transformar la base de datos y obtener resultados a nivel de gastos promedio por hogar a nivel de divisiones para el total de capitales regionales:

```
Stata
//Calcular el gasto promedio por hogar a nivel de divisiones para el total de
capitales regionales (Sin Arriendo Imputado)
cd "[directorio donde se encuentran las bases de datos]"
use "BASE_GASTOS_VIIEPF.dta", clear

//SE QUITA EL ARRIENDO IMPUTADO
drop if (CCIF=="04.2.1.01.01"|CCIF=="04.2.2.01.01"|CCIF=="04.2.2.01.02")

gen double GASTO_EXP=GASTO*FE //Gasto Expandido por hogar

//Generar número total de hogares
bysort FOLIO: gen AUX=_n==1
gen double AUX2=AUX*FE //Número de hogares que representa cada hogar
encuestado
egen double Total_hogares_VII=sum(AUX2) //Número total de hogares
encuestados

collapse (sum) GASTO_EXP, by(D Total_hogares_VII)

gen double GASTO_D= GASTO_EXP/ Total_hogares_VII //Variable con el gasto
promedio por hogar por división
g G="0"
g C="0"
g SC="00"
g P="00"
merge 1:m D G C SC P using "CCIF_VIIEPF.dta" //unión con la base CCIF para
obtener las glosas a nivel producto.
drop if _merge==2 //Eliminar los bienes y servicios del resto de las categorías.
Grupo-Clase-Subclase-Producto
drop _merge
```

Las variables G, C, SC y P son variables auxiliares que sirven para unir la base de datos con la base de datos donde se encuentran las glosas de las divisiones (primer nivel de desagregación). La base de datos de gasto incorpora las glosas a nivel de producto, por lo tanto si se quieren las glosas a otros niveles de desagregación, la base de datos debe ser trabajada.

Para obtener el mismo tabulado detallado anteriormente, pero esta vez incluyendo el arriendo imputado en el gasto de los hogares, se debe eliminar el

comando que borra los códigos de arriendo imputado (códigos="04.2.1.01.01", "04.2.2.01.01" y "04.2.2.01.02")

Stata

```
//Calcular el gasto promedio por hogar a nivel de divisiones para el total de
capitales regionales (Con Arriendo Imputado)
cd "[directorio donde se encuentran las bases de datos]"
use "BASE_GASTOS_VIIEPF.dta", clear

gen double GASTO_EXP=GASTO*FE //Gasto Expandido por hogar

//Generar número total de hogares
bysort FOLIO: gen AUX=_n==1
gen double AUX2=AUX*FE //Número de hogares que representa cada hogar
encuestado
egen double Total_hogares_VII=sum(AUX2) //Número total de hogares
encuestados

collapse (sum) GASTO_EXP, by(D Total_hogares_VII)

gen double GASTO_D= GASTO_EXP/ Total_hogares_VII //Variable con el gasto
promedio por hogar por división
g G="0"
g C="0"
g SC="00"
g P="00"
merge 1:m D G C SC P using "CCIF_VIIEPF.dta" //unión con la base CCIF para
obtener las glosas a nivel producto.
drop if _merge==2 //Eliminar los bienes y servicios del resto de las categorías.
Grupo-Clase-Subclase-Producto
drop _merge
```

Si se requiere un nivel más agregado de información, por ejemplo a nivel de grupos (que corresponde a la segunda desagregación de la CCIF), se debe colapsar la base agregando el nivel de agregación deseado. Al momento de crear las variables auxiliares para la unión con la base de glosas de CCIF (a través del comando merge) se debe tener cuidado de no crear la auxiliar según la cual se colapsó la base de datos.

Stata

```
//Calcular el gasto promedio por hogar a nivel de Grupos para el total de
capitales regionales (Sin Arriendo Imputado)
cd "[directorio donde se encuentran las bases de datos]"
use "BASE_GASTOS_VIIEPF.dta", clear

//SE QUITA EL ARRIENDO IMPUTADO
drop if (CCIF=="04.2.1.01.01"|CCIF=="04.2.2.01.01"|CCIF=="04.2.2.01.02")

gen double GASTO_EXP=GASTO*FE //Gasto Expandido por hogar

//Generar número total de hogares
bysort FOLIO: gen AUX=_n==1
gen double AUX2=AUX*FE //Número de hogares que representa cada hogar
encuestado
egen double Total_hogares_VII=sum(AUX2) //Número total de hogares
encuestados

collapse (sum) GASTO_EXP, by(D G Total_hogares_VII)

gen double GASTO_D= GASTO_EXP/ Total_hogares_VII //Variable con el gasto
promedio por hogar por división
g C="0"
g SC="00"
g P="00"
merge 1:m D G C SC P using "CCIF_VIIEPF.dta" //unión con la base CCIF para
obtener las glosas a nivel producto
drop if _merge==2 //Eliminar los bienes y servicios del resto de las categorías.
Grupo-Clase-Subclase-Producto
drop _merge
```

Nuevamente, si se quiere calcular el gasto para el total de capitales regionales a nivel de grupos, esta vez con el arriendo imputado, se debe aplicar la misma sintaxis antes descrita, eliminando el comando que borra los códigos de arriendo imputado.

Stata

```
//Calcular el gasto promedio por hogar a nivel de Grupos (Con Arriendo
Imputado)
cd "[directorio donde se encuentran las bases de datos]"
use "BASE_GASTOS_VIIEPF.dta", clear

gen double GASTO_EXP=GASTO*FE //Gasto Expandido por hogar

//Generar número total de hogares
bysort FOLIO: gen AUX=_n==1
gen double AUX2=AUX*FE //Número de hogares que representa cada hogar
encuestado
egen double Total_hogares_VII=sum(AUX2) //Número total de hogares
```

```
encuestados
```

```
collapse (sum) GASTO_EXP, by(D G Total_hogares_VII)
```

```
gen double GASTO_D= GASTO_EXP/ Total_hogares_VII //Variable con el gasto promedio por hogar por división
```

```
g C="0"
```

```
g SC="00"
```

```
g P="00"
```

```
merge 1:m D G C SC P using "CCIF_VIIEPF.dta" //unión con la base CCIF para obtener las glosas a nivel producto
```

```
drop if _merge==2 //Eliminar los bienes y servicios del resto de las categorías.
```

```
Grupo-Clase-Subclase-Producto
```

```
drop _merge
```

Si se requieren otros niveles de desagregación se deben repetir los pasos antes descritos incorporando el nivel de desagregación deseado en la programación.

- **Gasto total por hogar (con y sin arriendo imputado)**

Para el cálculo del gasto total por hogar es necesario sumar todos los gastos realizados por el hogar. Para hacer simultáneamente el cálculo del ingreso con y sin el arriendo imputado se identifican los códigos de arriendo imputado asociados a la primera y segunda vivienda⁷:

Stata

```
/*Recordar los códigos de arriendo imputado*/
```

```
use "BASE_GASTOS_VIIEPF", clear
```

```
gen double GASTOT_FNR= GASTO if (CCIF!="04.2.1.01.01" &  
CCIF!="04.2.2.01.01" & CCIF!="04.2.2.01.02")
```

```
gen double GASTOT_FNR_AI= GASTO
```

```
collapse (sum) GASTOT_FNR GASTOT_FNR_AI , by(FOLIO)
```

⁷ "04.2.1.01.01"-Arriendo imputado de la vivienda principal sólo para propietarios.

"04.2.2.01.01"- Arriendo imputado de la vivienda principal para hogares que pagan arriendo reducido o gratuito.

"04.2.2.01.02"- Arriendo imputado segunda vivienda.

4.2. Consolidación de las bases

Para realizar la unión entre las bases de personas y gastos, las personas deben ser ordenadas de acuerdo al identificador del hogar (FOLIO). Se debe tener en cuenta que la base de gasto se encuentra a nivel hogar y no a nivel persona, por lo que, para la unión de ambas bases se debe tener en cuenta dicha consideración y definir el nivel (hogar, jefe de hogar, etc.) al que se desea hacer el análisis.

A continuación se presenta un ejemplo de cómo puede realizarse la programación en Stata, para la unión de las bases de gastos e ingresos, tomando como nivel de análisis al jefe de hogar en la base de personas:

```
Stata
clear
use "BASE_GASTOS_VIIEPF.dta", clear
gen PERSONA=1 // Se genera la variable persona para asignar los
//gastos al jefe del hogar
sort FOLIO PERSONA
save "BASE_GASTOS_VIIEPF.dta", replace

preserve
use "BASE_PERSONAS_VIIEPF.dta", clear
keep if PERSONA==1
sort FOLIO PERSONA
save "BASE_AUXILIAR.dta", replace
restore

merge FOLIO PERSONA using "BASE_AUXILIAR.dta"
```

4.3. Quintilización por ingreso

Las bases publicadas no se presentan quintilizadas, sin embargo, en esta sección se explica cómo poder replicar las quintilizaciones utilizadas para la realización de los tabulados, no obstante, se debe tener en cuenta que los quintiles sólo corresponden a una variable de ordenación que indica en qué 20% de la distribución de los ingresos (del hogar o per cápita) se encuentra el hogar y no corresponde a una variable de estratificación socioeconómica, ya que para lograr una estratificación se deben incorporar al análisis otras dimensiones. Además, los análisis por quintiles tienen carácter descriptivo y no cuentan con significancia estadística.

Existen cuatro definiciones de ingresos por los cuales se realizan la ordenación y clasificación de los hogares de acuerdo al quintil al que pertenecen; estas son: Ingreso disponible total por hogar sin arriendo imputado, Ingreso disponible total por hogar con arriendo imputado, Ingreso disponible per cápita por hogar sin arriendo imputado, y finalmente, Ingreso disponible per cápita por hogar con arriendo imputado.

Por otro lado, cada una de estas definiciones se aplicaron para tres ámbitos geográficos: Nacional, Gran Santiago y Resto regiones. Como resultado, los tabulados presentados en el micro-sitio fueron realizados según doce quintilizaciones diferentes, que corresponden a las combinaciones entre las cuatro definiciones de la variable principal de ordenamiento y las tres desagregaciones geográficas.

Para la quintilización correcta de la base a nivel personas, es necesario considerar que al ordenar la base, se identificaran hogares que estén en el límite del corte entre un quintil y el siguiente. Es así, que para obtener el 20%⁸ exacto se debe particionar a cuatro hogares (para cada tipo de quintilización). Esto se

⁸ El proceso de quintilización obtiene 5 grupos con la misma cantidad de hogares cada uno, representando cada uno de estos grupos un 20% de los hogares respecto al total nacional. Estos grupos son obtenidos a partir de una determinada ordenación realizada a través de los ingresos según se explicó con anterioridad. El primero de estos grupos contiene al 20% de los hogares de menos ingresos, mientras que el quinto quintil aglutina el 20% de los hogares más ricos según la variable de ordenación.

realiza sobre la variable del factor de expansión, ya que las demás características se mantienen sin variación.

Stata

```
/*EJEMPLO: Ámbito: Nacional */
/*Variable de ordenamiento: Ingreso disponible total por hogar sin arriendo
imputado */

use "BASE_PERSONAS_VIIEPF.dta", clear
replace INGDHOG_HD=. if INGDHOG_HD <=0
replace AI_VP=0 if AI_VP==.
replace AI_VS=0 if AI_VS==.

tabstat INGDHOG_HD if PERSONA==1, stat(p1)
gen double id_missing=1 if INGDHOG_HD ==. | INGDHOG_HD <=70000
replace INGDHOG_HD =GASTOT_FNR if id_missing==1
replace INGDHOG_HD_AI=GASTOT_FNR+ AI_VP+AI_VS if id_missing==1

foreach v in INGA_HD INGH_HD INGCP_HD INGPI_HD INGJ_HD INGOTA INGOTI
{
  replace `v'=0 if `v'==.
}
gen double
ORDEN_B=INGA_HD+INGH_HD+INGCP_HD+INGPI_HD+INGJ_HD+INGOTA+INGOTI
replace ORDEN_B =. if ORDEN_B ==0
by FOLIO, sort: egen double ORDEN_BH=total(ORDEN_B)
// NIVEL HOGAR
keep if PERSONA==1
gen double INGHOG_HD_PC=INGHOG_HD/NPERSONA
gen double INGDHOG_HD_PC= INGDHOG_HD /NPERSONA
gen double GASTOT_FNR_PC=GASTOT_FNR/ NPERSONA
//ordenamiento de los hogares
sort INGDHOG_HD INGDHOG_HD_PC ORDEN_BH GASTOT_FNR FOLIO
g double s=sum(FE)
g double c=(s)/(s[_N])
g double tag_QNSAI=0
g double w2=.
qui foreach l in 20 40 60 80 {
  replace tag_QNSAI=1 if c>=`l'/100 & c[_n-1]<`l'/100
  replace w2=s-(`l'/100*s[_N]) if c>=`l'/100 & c[_n-1]<`l'/100
}
g double w1=FE-w2
g double id=_n
expand 2 if w2<.
g double FW_QNSAI=FE
bys id: replace FW_QNSAI=w1 if _n==1 & w1<.
by id: replace FW_QNSAI=w2 if _n==2
g double sum=sum(FW_QNSAI)
g double cum=sum/sum[_N]
```

```

g q_NSAI=1 if cum<=0.2
replace q_NSAI=2 if cum<=0.4 & cum>0.2
replace q_NSAI=3 if cum<=0.6 & cum>0.4
replace q_NSAI=4 if cum<=0.8 & cum>0.6
replace q_NSAI=5 if cum<=1 & cum>0.8
list FOLIO FE FW_QNSAI q_NSAI INGHOG_HD INGDHOG_HD_PC ORDEN_BH
GASTOT_FNR if tag_QNSAI==1
keep FOLIO PERSONA FE FW_QNSAI q_NSAI tag_QNSAI ID_MISSING
sort FOLIO PERSONA
// se guarda una base auxiliar a nivel hogar
bysort FOLIO: gen aux=_n-1 //se crea una variable auxiliar que será utilizada
para unir con la base de gasto
save "BASE_QNSAI_HOGAR.dta", replace

use "BASE_PERSONAS_VIIEPF.dta", clear
sort FOLIO PERSONA
merge FOLIO PERSONA using "BASE_QNSAI_HOGAR.dta", force
replace _merge=. if _merge!=3
by FOLIO,sort: egen sum_merge=total(_merge)
tab sum_merge
expand 2 if sum_merge==6 & PERSONA!=1
bys FOLIO PERSONA: gen id=_n
bys FOLIO id: egen double FW_QNSAIH=total(FW_QNSAI)
bys FOLIO id: egen double q_NSAIH=total(q_NSAI)
bys FOLIO id: egen double tag_QNSAIH=total(tag_QNSAI)
drop FW_QNSAI q_NSAI _merge sum_merge id tag_QNSAI
ren FW_QNSAIH FW_QNSAI
ren q_NSAIH q_NSAI
ren tag_QNSAIH tag_QNSAI
tabstat FW_QNSAI if PERSONA==1, stat(sum) by(q_NSAI) format(%12.0f)

keep FOLIO PERSONA FW_QNSAI q_NSAI
save "BASE_QNSAI_PERSONAS.dta", replace

//El comando para ordenar a los hogares para cada tipo de quintilización  

restante sigue la siguiente lógica (Estos comandos pueden ser  

reemplazados en la línea con negrita para obtener distintas  

quintilizaciones):
/* Ingreso disponible total del hogar sin arriendo imputado */
sort INGHOG_HD ING_DISP_HOGAR_PC_HD ORDEN_BH GASTOT_FNR FOLIO
/* Ingreso disponible total del hogar con arriendo imputado */
sort INGHOG_HD_AI ING_DISP_HOGAR_PC_HD_AI ORDEN_BH GASTOT_FNR_AI
FOLIO
/* Ingreso disponible total del hogar per cápita sin arriendo imputado */
sort ING_DISP_HOGAR_PC_HD ORDEN_BH_PC_HD INGHOG_HD
GASTOT_FNR_PC FOLIO
/* Ingreso disponible total del hogar per cápita con arriendo imputado */
sort ING_DISP_HOGAR_PC_HD_AI ORDEN_BH_PC_HD GASTOT_FNR_AI_PC
INGHOG_HD FOLIO

```


Tabulado de la composición del ingreso de los grupos quintiles de hogares ordenados de acuerdo al ingreso del hogar, según fuentes de ingresos, total nacional (excluye arriendo imputado)

Stata

```
use "BASE_PERSONAS_VIIEPF.dta", clear
quietly merge 1:m FOLIO PERSONA using "BASE_QNSAI_PERSONAS.dta",
keepusing(FW_QNSAI q_NSAI) generate(MERGE)
tabstat FW_QNSAI if PERSONA==1, stat(sum) by (q_NSAI) format(%22.5f)
foreach N in INGDA_HD INGDH_HD INGDCP_HD INGDPI_HD INGOTA INGOTI
INGP INGJ_HD INGT_NETA {
quietly replace `N'=0 if `N'==. | INGDHOG_HD==.
}
foreach N in AI_VP AI_VS2 {
quietly replace `N'=0 if `N'==.
}
quietly gen double DE_LA_OCUPACION=INGDA_HD + INGDH_HD + INGDCP_HD
+ INGDPI_HD + INGOTA + INGOTI
quietly gen double SUELDOS_Y_SALARIOS=INGDA_HD + INGDH_HD
quietly gen double ING_I=INGDCP_HD + INGDPI_HD
quietly gen double EMPLEADORES=ING_I if AECISE==1
quietly gen double CUENTA_PROPIA=ING_I if AECISE==2
quietly gen double OTROS_TRABAJOS=INGOTA + INGOTI
quietly gen double RENTAS_PROPIEDAD=INGP + INGP
quietly gen double TRANSFERENCIAS=INGT_NETA
quietly gen double ARRIENDO_IMPUTADO=AI_VP+AI_VS2
    quietly replace ARRIENDO_IMPUTADO=0 if PERSONA!=1
quietly gen double JUBILACION=INGJ_HD*0.93
quietly gen double
OTROS_ING=JUBILACION+RENTAS_PROPIEDAD+TRANSFERENCIAS
quietly gen double
OTROS_ING_AI=JUBILACION+RENTAS_PROPIEDAD+ARRIENDO_IMPUTADO+TR
ANSFERENCIAS
quietly gen double
TOTAL=DE_LA_OCUPACION+JUBILACION+RENTAS_PROPIEDAD+TRANSFERENCI
AS
quietly gen double
TOTAL_AI=DE_LA_OCUPACION+JUBILACION+RENTAS_PROPIEDAD+ARRIENDO_
IMPUTADO+TRANSFERENCIAS
foreach v in TOTAL TOTAL_AI DE_LA_OCUPACION SUELDOS_Y_SALARIOS ING_I
EMPLEADORES CUENTA_PROPIA OTROS_TRABAJOS OTROS_ING
OTROS_ING_AI JUBILACION RENTAS_PROPIEDAD ARRIENDO_IMPUTADO AI_VP
AI_VS2 TRANSFERENCIAS {
quietly bys FOLIO q_NSAI: egen double `v'_H=total(`v')
quietly replace `v'_H=0 if PERSONA!=1
}
```

```

foreach v in TOTAL TOTAL_AI DE_LA_OCUPACION SUELDOS_Y_SALARIOS ING_I
EMPLEADORES CUENTA_PROPIA OTROS_TRABAJOS OTROS_ING
OTROS_ING_AI JUBILACION RENTAS_PROPIEDAD ARRIENDO_IMPUTADO AI_VP
AI_VS2 TRANSFERENCIAS {
quietly gen double `v'_EXP=`v'_H*FW_QNSAI
}
***** PARA OBTENER LOS PROMEDIOS DIRECTAMENTE *****
quietly egen double NHOG_TOT=sum(FW_QNSAI) if PERSONA==1
quietly gen double NHOG_q=NHOG_TOT/5
foreach ING in TOTAL_EXP DE_LA_OCUPACION_EXP SUELDOS_Y_SALARIOS_EXP
EMPLEADORES_EXP CUENTA_PROPIA_EXP OTROS_TRABAJOS_EXP
OTROS_ING_EXP JUBILACION_EXP RENTAS_PROPIEDAD_EXP
TRANSFERENCIAS_EXP {
foreach q in 1 2 3 4 5 {
quietly egen double `ING'_'q'=sum(`ING') if q_NSAI==`q'
quietly replace `ING'_'q'=(`ING'_'q'/NHOG_q)
}
}
collapse (sum) TOTAL_EXP DE_LA_OCUPACION_EXP
SUELDOS_Y_SALARIOS_EXP EMPLEADORES_EXP CUENTA_PROPIA_EXP
OTROS_TRABAJOS_EXP OTROS_ING_EXP JUBILACION_EXP
RENTAS_PROPIEDAD_EXP TRANSFERENCIAS_EXP (mean) NHOG_q,by(q_NSAI)

foreach v in TOTAL_EXP DE_LA_OCUPACION_EXP SUELDOS_Y_SALARIOS_EXP
EMPLEADORES_EXP CUENTA_PROPIA_EXP OTROS_TRABAJOS_EXP
OTROS_ING_EXP JUBILACION_EXP RENTAS_PROPIEDAD_EXP
TRANSFERENCIAS_EXP {
replace `v'=`v'/ NHOG_q
}
//PARA OBTENER LOS RESULTADOS TRANSPUESTOS
sxpose, clear force firstnames destring
foreach x in 1 2 3 4 5 {
ren _var`x' q_`x'
}
gen FUENTES="TOTAL" if _n==1
replace FUENTES="DE_LA_OCUPACION" if _n==2
replace FUENTES="SUELDOS_Y_SALARIOS" if _n==3
replace FUENTES="EMPLEADORES" if _n==4
replace FUENTES="CUENTA_PROPIA" if _n==5
replace FUENTES="OTROS_TRABAJOS" if _n==6
replace FUENTES="OTROS_ING" if _n==7
replace FUENTES="JUBILACION" if _n==8
replace FUENTES="RENTAS_PROPIEDAD" if _n==9
replace FUENTES="TRANSFERENCIAS" if _n==10
replace FUENTES="N° de Hogares" if _n==11
order FUENTES

```

- **Gasto total promedio por hogar por división y quintil**

A continuación se detalla cómo obtener el gasto promedio del hogar para cada una de las divisiones que conforman la CCIF. Para utilizar estos comandos, es necesario haber ejecutado los comandos de quintilización antes descritos.

Stata

```
use "BASE_GASTOS_VIIEPF.dta", clear

expand 2 if FOLIO=="02282-1"|FOLIO=="09764-1"|FOLIO=="14587-1"|FOLIO=="17289-2", generate (aux) //la expansión de los hogares de corte, dependerá del tipo de método de quintilización utilizado
drop if GASTO==0 //Eliminamos los códigos sin gasto

//SE QUITA EL ARRIENDO IMPUTADO
drop if (CCIF=="04.2.1.01.01"|CCIF=="04.2.2.01.01"|CCIF=="04.2.2.01.02")
//merge m:1 DIR_SUB_MUESTRA FOLIO using "C:\EPF\DATA\LEVANT\VII EPF\BASE FINAL\BASES AUX\FE.dta"
merge m:m FOLIO aux using "BASE_QNSAI_HOGAR.dta" //SOLO DEBE EXISTIR _MERGE==3
drop _merge

gen double gasto_exp = GASTO*FW_QNSAI
replace gasto_exp=0 if gasto_exp==.

by FOLIO q_NSAI, sort: gen AUX=_n==1
gen n_hogares=AUX* FW_QNSAI
bysort q_NSAI: egen double n_hogares_totales=sum(n_hogares) //Número total de hogares encuestados por quintil
bysort q_NSAI: egen double gasto_total_hogares=sum(gasto_exp) //Gasto total de todos los hogares por quintil
bysort q_NSAI: gen double
G_Total_HOG=gasto_total_hogares/n_hogares_totales

collapse (sum) gasto_exp, by (CCIF GLOSA G_Total_HOG q_NSAI n_hogares_totales)

gen double Gasto_hog_prod= gasto_exp/ n_hogares_totales
gen double Porcentaje_hog_prod= (Gasto_hog_prod/ G_Total_HOG)*100

/* //Hasta aquí ya está la base a nivel de gasto promedio por hogar en cada producto para cada quintil
//Para guardar la base, se pueden aplicar los siguientes comandos:
gen _mi_miss=0
mi unset
drop mi_miss n_hogares_totales G_Total_HOG gasto_exp
reshape wide Gasto_hog_prod Porcentaje_hog_prod, i( CCIF GLOSA) j( q_NSAI)
save "BASE_q_Productos.dta"
*/
```

```

*****A NIVEL DE DIVISIONES *****
gen D=substr(CCIF,1,2)
collapse(sum) Gasto_hog_prod, by (D q_NSAI G_Total_HOG)
gen double Porcentaje_hog_prod = (Gasto_hog_prod/ G_Total_HOG)*100
sort q_NSAI D
order q_NSAI D Gasto_hog_prod Porcentaje_hog_prod G_Total_HOG
gen G="0"
gen C="0"
gen SC="00"
gen P="00"
merge m:m D G C SC P using "CCIF_VIIEPF.dta" //No debe existir merge==1
drop if _merge==2
order q_NSAI CCIF GLOSA Gasto_hog_prod Porcentaje_hog_prod
keep q_NSAI CCIF GLOSA Gasto_hog_prod Porcentaje_hog_prod
//PARA APLICAR reshape
gen _mi_miss=0
mi unset
drop mi_miss
reshape wide Gasto_hog_prod Porcentaje_hog_prod , i( CCIF GLOSA) j(q_NSAI)

```

La programación detalla cómo obtener el gasto promedio por hogar y la estructura del gasto promedio por hogar para cada producto y luego en una segunda etapa, para cada división. Si se requiere otro nivel de agregación (grupos, clases, subclases) se debe modificar la segunda parte de la programación.