



PUCP

STATA¹⁷

STATISTICS • VISUALIZATION • DATA MANIPULATION • REPORTING

DEPARTAMENTO DE ECONOMÍA
LABORATORIO DE ECONOMETRÍA: STATA
1ECO31

Sesión 2

Manipulación de Datos con ENAHO

Docente: Juan Palomino



PUCP

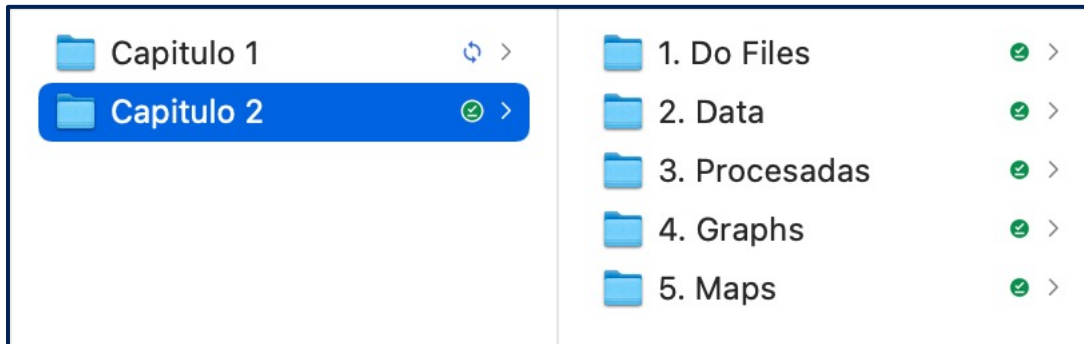
Índice

| | | | | | |
|---|-------------------------------------|---|-----------------------------|----|-----------------------------------|
| 1 | Organización del Proyecto | 5 | Explorando Variables | 9 | Borrar y Mantener Variables |
| 2 | Descargando Base de Datos de la Web | 6 | Tipo y Formato de Variables | 10 | Ordenar Observaciones y Variables |
| 3 | Cargando Base de Datos | 7 | Manipulando Variables | 11 | Etiquetas |
| 4 | Traducir Base de Datos | 8 | Manipulando Observaciones | 12 | Guardar y Exportar |

1. Organización del Proyecto

¡Aprender a cómo organizar nuestros archivos de trabajo!.

Organizar información en carpetas



1. **Do Files:** se guardan los do files de las programaciones.
2. **Data:** se guardan los archivos original de la data.
3. **Procesadas:** se guardan las bases de datos procesadas a partir de los archivos originales.
4. **Graphs:** se guardan las figuras en png, gph, jpg.
5. **Maps:** se guardan los archivos shapefiles.

Criterios de Organización

Una vez que hemos organizado nuestras carpetas, fijaremos la carpeta de trabajo donde estará nuestro proyecto actual con el comando “**cd**”:

```
* Primera Opción  
cd "/Users/juanpalomino/Google Drive/ENAH0/Sumaria"
```

También se puede guardar a través del comando “**global**” las rutas de las carpetas, y usarlas en los momentos apropiados:

```
* Segunda Opción  
global main "/Users/juanpalomino/Google Drive/Cursos/Laboratorio de Stata"  
global dos "$main/1. Do Files"  
global dta "$main/2. Data"  
global works "$main/3. Procesadas"
```

2. Descargando de la Web

Descargando Base de Datos – Primera Forma

<http://iinei.inei.gov.pe/microdatos/>

Pasos:

1. Típear www.inei.gov.pe
2. Clic en Bases de Datos y luego en Microdatos
3. Clic en Consulta por Encuestas
4. Escoger Enaho Metodología Actualizada
5. Elegir el año, periodo y el modulo de interés








































CONSULTA POR ENCUESTA

Sírvase seleccionar Encuesta, Año y Periodo y a continuación se mostrarán todas los Módulos de la Encuesta Seleccionada. Luego proceda a descargar el módulo de su interés.

ENCUESTA: **ENAHO Metodología ACTUALIZADA**

Condiciones de Vida y Pobreza - ENAHO

AÑO: 2020 Periodo: Anual - (Ene-Dic)

| Nro | Año | Periodo | Código Encuesta | Encuesta | Código Módulo | Módulo | Ficha | Descarga |
|-----|------|---------|-----------------|---------------------------------------|---------------|-------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| 1 | 2020 | 55 | 737 | Condiciones de Vida y Pobreza - ENAHO | 1 | Características de la Vivienda y del Hogar |   |  |
| 2 | 2020 | 55 | 737 | Condiciones de Vida y Pobreza - ENAHO | 2 | Características de los Miembros del Hogar |   |  |
| 3 | 2020 | 55 | 737 | Condiciones de Vida y Pobreza - ENAHO | 3 | Educación |   |  |
| 4 | 2020 | 55 | 737 | Condiciones de Vida y Pobreza - ENAHO | 4 | Salud |   |  |
| 5 | 2020 | 55 | 737 | Condiciones de Vida y Pobreza - ENAHO | 5 | Empleo e Ingresos |   |  |
| 6 | 2020 | 55 | 737 | Condiciones de Vida y Pobreza - ENAHO | 7 | Gastos en Alimentos y Bebidas (Módulo 601) |   |  |
| 7 | 2020 | 55 | 737 | Condiciones de Vida y Pobreza - ENAHO | 8 | Instituciones Benéficas |   |  |
| 8 | 2020 | 55 | 737 | Condiciones de Vida y Pobreza - ENAHO | 9 | Mantenimiento de la Vivienda |   |  |
| 9 | 2020 | 55 | 737 | Condiciones de Vida y Pobreza - ENAHO | 10 | Transportes y Comunicaciones |   |  |
| 10 | 2020 | 55 | 737 | Condiciones de Vida y Pobreza - ENAHO | 11 | Servicios a la Vivienda |   |  |
| 11 | 2020 | 55 | 737 | Condiciones de Vida y Pobreza - ENAHO | 12 | Esparcimiento, Diversión y Servicios de Cultura |   |  |
| 12 | 2020 | 55 | 737 | Condiciones de Vida y Pobreza - ENAHO | 13 | Vestido y Calzado |   |  |
| 13 | 2020 | 55 | 737 | Condiciones de Vida y Pobreza - ENAHO | 15 | Gastos de Transferencias |   |  |

Descargando Base de Datos – Segunda Forma

1. Pasar el mouse por el archivo zip de la encuesta a descargar y aparecerá un link



















CONSULTA POR ENCUESTA

Sírvase seleccionar Encuesta, Año y Período y a continuación se mostrarán todas los Módulos de la Encuesta Seleccionada. Luego proceda a descargar el módulo de su interés.

ENCUESTA: ENAHO Metodología ACTUALIZADA


































Condiciones de Vida y Pobreza - ENAHO

AÑO: 2021 Período: Anual - (Ene-Dic)

| Nro | Año | Período | Código Encuesta | Encuesta | Código Módulo | Módulo | Ficha | Descarga |
|-----|------|---------|-----------------|---------------------------------------|---------------|--------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| 1 | 2021 | 55 | 759 | Condiciones de Vida y Pobreza - ENAHO | 1 | Características de la Vivienda y del Hogar |    | CSV |
| 2 | 2021 | 55 | 759 | Condiciones de Vida y Pobreza - ENAHO | 2 | Características de los Miembros del Hogar |    | CSV |
| 3 | 2021 | 55 | 759 | Condiciones de Vida y Pobreza - ENAHO | 3 | Educación |    | CSV |
| 4 | 2021 | 55 | 759 | Condiciones de Vida y Pobreza - ENAHO | 4 | Salud |    | CSV |
| 5 | 2021 | 55 | 759 | Condiciones de Vida y Pobreza - ENAHO | 5 | Empleo e Ingresos |    | CSV |
| | | | | Condiciones de Vida y Pobreza - ENAHO | 7 | Gastos en Alimentos y Bebidas (Módulo 601) |    | CSV |

inei.inei.gob.pe/inei/srienaho/descarga/STATA/759-Modulo05.zip

2. Clic derecho en el zip y copiar el link

| Código Módulo | Módulo | Ficha | Descarga |
|---------------|--------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| 1 | Características de la Vivienda y del Hogar |    | CSV |
| 2 | Características de los Miembros del Hogar |    | CSV |
| 3 | Educación |    | CSV |
| 4 | Salud |    | CSV |
| 5 | Em |    | CSV |
| 7 | Gas (Módulo 601) |    | CSV |
| 8 | Ins |    | CSV |
| 9 | Ma |    | CSV |
| 10 | Tra |    | CSV |
| 11 | Ser |    | CSV |
| 12 | Esp de |    | CSV |

Descargando Base de Datos – Segunda Forma

3. El siguiente código te permitira descargar la carpeta de la encuesta:

```
*=====
* II. Descargando de la web una base de datos
*=====

global inei "http://iinei.inei.gob.pe/iinei/srienaho/descarga/STATA/"
copy "$inei/759-Modulo05.zip" "759-Modulo05.zip", replace
unzipfile "759-Modulo05", replace
erase "759-Modulo05.zip"
```

¿Qué hace?

- Copia el archivo zip de la encuesta a tu carpeta local o al CD especificado anteriormente
- Con unzipfile descomprime el zip de la encuesta
- Con erase borra el archivo zip de la encuesta

2. Cargando Base de Datos

Conceptualización de los Datos

Los datos deben estar presentados tal que las columnas sean las variables y las filas sus observaciones.

| | | Variables | | | | | | | |
|---------------|----|-----------|----------|-------|--------|----------|-----------|-----------|----------|
| Observaciones | | conglome | vivienda | hogar | ubigeo | mieperho | inghog2d | gashog2d | pobreza |
| | 1 | 005009 | 012 | 11 | 010101 | 2 | 37274.426 | 25065.797 | no pobre |
| | 2 | 005009 | 041 | 11 | 010101 | 2 | 41967.57 | 28650.02 | no pobre |
| | 3 | 005009 | 098 | 11 | 010101 | 6 | 42292.535 | 38136.219 | no pobre |
| | 4 | 005015 | 007 | 11 | 010101 | 4 | 37866.211 | 26812.154 | no pobre |
| | 5 | 005015 | 019 | 11 | 010101 | 4 | 91681.961 | 63311.539 | no pobre |
| | 6 | 005015 | 070 | 11 | 010101 | 4 | 29749.256 | 59986.023 | no pobre |
| | 7 | 005015 | 085 | 11 | 010101 | 4 | 40505.797 | 33310.414 | no pobre |
| | 8 | 005015 | 117 | 11 | 010101 | 1 | 27953.023 | 21437.623 | no pobre |
| | 9 | 005032 | 014 | 11 | 010201 | 4 | 77708.43 | 38951.035 | no pobre |
| | 10 | 005032 | 047 | 11 | 010201 | 2 | 8458.1504 | 9408.3535 | no pobre |

Cargando Base de Datos

Stata tiene bases de datos predeterminadas en su memoria y estas se pueden abrir con el comando “`sysuse`”

```
* El comando sysuse  
sysuse auto, clear
```

Si se quiere abrir un archivo de excel se usa “`import excel`”:

```
* El comando import excel  
import excel "pwt100.xlsx", sheet("Data") firstrow clear
```

Para abrir una base de datos se usa el comando “`use`”:

```
* El comando use  
use "enaho01a-2020-500.dta", clear
```

4. Traducir Base de Datos

Traducir Base de Datos

Problema: unicode o caracteres con signo de interrogación que no permiten leer adecuadamente la base de datos

Solución: usar el comando unicode para traducir la base de datos, pero para eso no se debe tener base abierta en Stata

| año[1] | | 2020 | | | |
|--------|------|----------|------------|------------|------------|
| | año | codperso | estrato | p558a5 | p558d2_1 |
| 1 | 2020 | 01 | de 20 0... | no est0... | aqu0, e... |
| 2 | 2020 | 02 | de 20 0... | no est0... | aqu0, e... |
| 3 | 2020 | 01 | de 20 0... | no est0... | aqu0, e... |
| 4 | 2020 | 02 | de 20 0... | no est0... | aqu0, e... |
| 5 | 2020 | 01 | de 20 0... | no est0... | . |
| 6 | 2020 | 02 | de 20 0... | no est0... | aqu0, e... |
| 7 | 2020 | 03 | de 20 0... | no est0... | aqu0, e... |
| 8 | 2020 | 04 | de 20 0... | no est0... | . |
| 9 | 2020 | 05 | de 20 0... | no est0... | . |
| 10 | 2020 | 01 | de 20 0... | no est0... | aqu0, e... |
| 11 | 2020 | 02 | de 20 0... | pase | aqu0, e... |
| 12 | 2020 | 03 | de 20 0... | no est0... | . |
| 13 | 2020 | 01 | de 20 0... | pase | . |

Variables

| | Name | Label |
|-------------------------------------|----------|------------------|
| <input checked="" type="checkbox"/> | año | año de la en... |
| <input type="checkbox"/> | mes | mes de ejecu... |
| <input type="checkbox"/> | conglome | número de c... |
| <input type="checkbox"/> | vivienda | número de s... |
| <input type="checkbox"/> | hogar | número sec... |
| <input checked="" type="checkbox"/> | codperso | número de o... |
| <input type="checkbox"/> | ubigeo | ubicación g... |
| <input type="checkbox"/> | dominio | dominio geo... |
| <input checked="" type="checkbox"/> | estrato | estrato geogr... |
| <input type="checkbox"/> | codinfor | código de in... |

```
clear all
unicode analyze "enaho01a-2020-500.dta"
unicode encoding set "latin1"
unicode translate "enaho01a-2020-500.dta"
```

Traducir Base de Datos

- **Primer Paso:** “unicode analyze” es para analizar si la base de datos contiene unicodes que se deben traducir a otro lenguaje.
- **Segundo Paso:** “unicode encoding set” es para establecer el lenguaje al cual se quiere traducir.
- **Tercer Paso:** “unicode translate” es para traducir la base de datos.

| año | codperso | estrato | p558a5 | p558d2_1 |
|------|----------|----------------------|------------------|------------------------|
| 2020 | 01 | Área de empadrona... | pase | aquí, en este distrito |
| 2020 | 02 | Área de empadrona... | no está afiliado | . |
| 2020 | 03 | Área de empadrona... | pase | en otro distrito |
| 2020 | 04 | Área de empadrona... | no está afiliado | . |
| 2020 | 01 | Área de empadrona... | pase | en otro distrito |
| 2020 | 02 | Área de empadrona... | no está afiliado | aquí, en este distrito |
| 2020 | 03 | Área de empadrona... | no está afiliado | . |
| 2020 | 04 | Área de empadrona... | no está afiliado | . |
| 2020 | 05 | Área de empadrona... | no está afiliado | aquí, en este distrito |

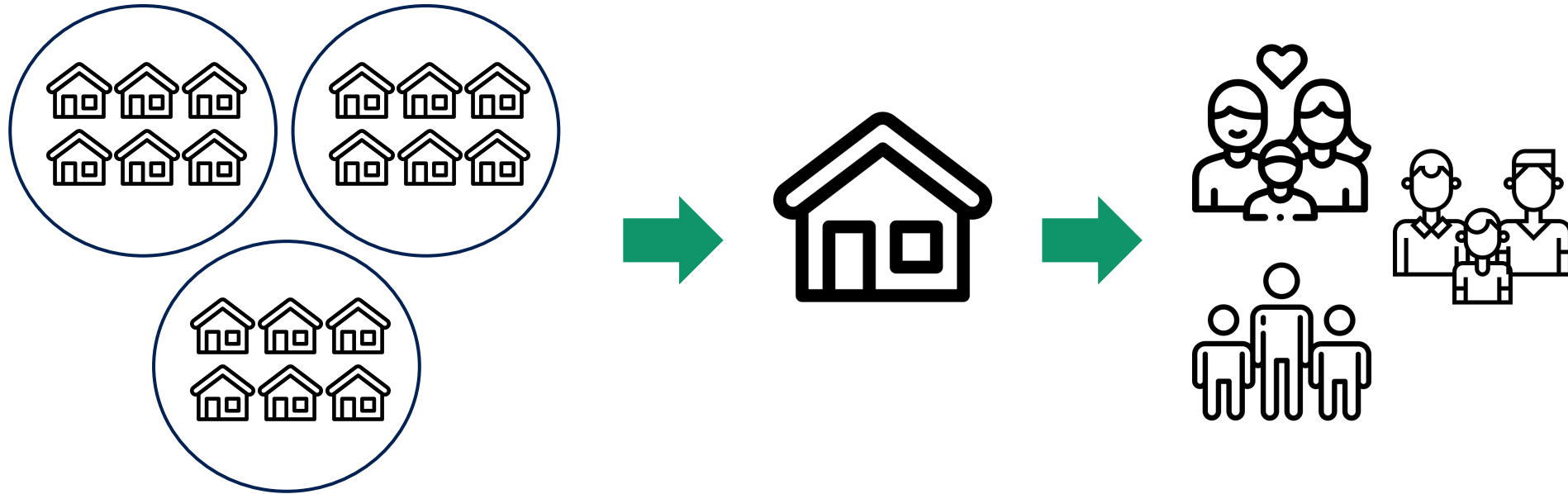
Variables

| | Name | Label |
|-------------------------------------|----------|-----------------|
| <input checked="" type="checkbox"/> | año | año de la en... |
| <input checked="" type="checkbox"/> | codperso | número de... |
| <input checked="" type="checkbox"/> | estrato | estrato geo... |
| <input checked="" type="checkbox"/> | p558a5 | ¿el sistema... |
| <input checked="" type="checkbox"/> | p558d2_1 | el lugar don... |
| <input type="checkbox"/> | mes | mes de ejec... |

5. Explorando Variables

Identificadores de la ENAHO

En la ENAHO hay bases que están a nivel de hogar y otras que están a nivel individual. Cada una de ellas tienen un identificador en común.



Id Hogar = Conglomerado



Vivienda



Hogar

Id Individual = Conglomerado



Vivienda



Hogar



Individuo

Comandos para explorar variables

Comando “**browse**”: sirve para visualizar las variables de interés en la base de datos.

```
* El comando browse  
browse  
browse ubigeo estrato conglome vivienda hogar codperso
```

Comando “**describe**”: sirve para describir el formato de las variables (nombre, tipo, formato, etiqueta).

```
* El comando describe  
describe  
describe ubigeo estrato conglome vivienda hogar codperso p207 p208a
```

Comando “**codebook**”: describe el contenido de la variable

```
* El comando codebook  
codebook p208a  
codebook ubigeo
```

6. Tipo y Formato de Variables

Tipo y Formato de Variables

Tipo de Variable: String

| conglome | vivienda | hogar | codperso | ubigeo |
|----------|----------|-------|----------|--------|
| 005009 | 012 | 11 | 01 | 010101 |
| 005009 | 012 | 11 | 02 | 010101 |
| 005009 | 041 | 11 | 01 | 010101 |
| 005009 | 041 | 11 | 02 | 010101 |
| 005009 | 098 | 11 | 01 | 010101 |
| 005009 | 098 | 11 | 02 | 010101 |
| 005009 | 098 | 11 | 03 | 010101 |
| 005009 | 098 | 11 | 04 | 010101 |
| 005009 | 098 | 11 | 05 | 010101 |
| 005015 | 007 | 11 | 01 | 010101 |

Formato Variable String

```
format ubigeo %6s
```

```
format ubigeo %18s
```

| ubigeo |
|--------|
| 150114 |
| 160101 |
| 180101 |
| 150108 |
| 150130 |
| 150136 |
| 060101 |
| 150121 |

| ubigeo |
|--------|
| 010101 |
| 010201 |
| 010701 |
| 010502 |
| 010501 |
| 010203 |
| 010306 |
| 010304 |

Tipo y Formato de Variables

Tipo de Variable: Numérica

| p207 | p208a | p524a1 |
|--------|-------|--------|
| mujer | 49 | . |
| mujer | 16 | . |
| mujer | 61 | . |
| mujer | 29 | 200 |
| hombre | 59 | . |
| mujer | 42 | . |
| mujer | 20 | 1200 |
| hombre | 18 | . |
| mujer | 16 | . |
| mujer | 36 | . |

Variable Numérica con
etiqueta

Formato Variable Numérica

```
format p524a1 %5.0f
```

| p524a1 |
|--------|
| 10000 |
| 10100 |
| 10500 |
| 10800 |
| 11000 |
| 12000 |
| 12500 |
| 12800 |

```
format p524a1 %5.2f
```

| p524a1 |
|----------|
| 10000.00 |
| 10100.00 |
| 10500.00 |
| 10800.00 |
| 11000.00 |
| 12000.00 |
| 12500.00 |
| 12800.00 |

```
format p524a1 %6.0fc
```

| p524a1 |
|--------|
| 10,000 |
| 10,100 |
| 10,500 |
| 10,800 |
| 11,000 |
| 12,000 |
| 12,500 |
| 12,800 |

Tipo y Formato de Variables

Variable Numérica con etiqueta

| estrato | p207 |
|-----------------------------------------------|--------|
| de 500 a 1 999 habitantes | mujer |
| de 500 a 1 999 habitantes | hombre |
| de 500 a 1 999 habitantes | hombre |
| de 500 a 1 999 habitantes | hombre |
| de 500 a 1 999 habitantes | mujer |
| de 500 a 1 999 habitantes | mujer |
| de 500 a 1 999 habitantes | mujer |
| Área de empadronamiento rural (aer) compuesto | hombre |
| Área de empadronamiento rural (aer) compuesto | mujer |
| Área de empadronamiento rural (aer) compuesto | mujer |
| Área de empadronamiento rural (aer) compuesto | mujer |
| Área de empadronamiento rural (aer) compuesto | hombre |
| Área de empadronamiento rural (aer) compuesto | hombre |

Para saber cual es el número que le corresponde a cada etiqueta de la variable se usa el comando “**label list**”

```
. label list p207
p207:
      1 hombre
      2 mujer
```

```
. label list estrato
estrato:
      1 de 500 000 a más habitantes
      2 de 100 000 a 499 999 habitantes
      3 de 50 000 a 99 999 habitantes
      4 de 20 000 a 49 999 habitantes
      5 de 2 000 a 19 999 habitantes
      6 de 500 a 1 999 habitantes
      7 Área de empadronamiento rural (aer) compuesto
      8 Área de empadronamiento rural (aer) simple
```

Missing Values

Son variables numéricas con algunas observaciones que no tienen valor.

| p524a1 | newvar |
|--------|--------|
| . | . |
| . | . |
| . | . |
| 200 | . |
| . | . |
| . | . |
| 1200 | . |
| . | . |
| . | . |
| . | . |
| . | . |

El comando “**mdesc**” permite saber la frecuencia y el porcentaje de missing que existen en una variable.

```
* Missing Values  
gen newvar=.  
ssc install mdesc  
mdesc p524a1 newvar
```



| Variable | Missing | Total | Percent Missing |
|----------|---------|--------|-----------------|
| p524a1 | 70,290 | 91,315 | 76.98 |
| newvar | 91,315 | 91,315 | 100.00 |

Formato de Variables

| Tipo | Almacenaje | Variable cuyo contenido es... | bytes |
|----------|------------|------------------------------------------------------|-------|
| Missing | missing | No hay valor | 1 |
| String | str1 | string de 1 carácter como máxima longitud | 1 |
| | str2 | string de 2 caracteres como máxima longitud | 2 |
| | str3 | string de 3 caracteres como máxima longitud | 3 |
| | ... | | |
| | strn | string de n caracteres como máxima longitud | n |
| Numérica | byte | números enteros entre -127 y 100 | 1 |
| | int | números enteros entre -32,767 y 32,740 | 2 |
| | long | números enteros entre -2,147,483,647 y 2,147,482,620 | 4 |
| | float | números reales con 8 dígitos de precisión | 4 |
| | double | números reales con 17 dígitos de precisión | 8 |

En el caso de los string, n debe ser menor o igual a 244.

7. Manipulando Variables

Renombrando Variables

El comando “**rename**” permite cambiar de nombre a las variables.

- Renombrando la variable edad:

```
* Edad  
rename p208a age
```

- Renombrando la variable sexo:

```
* Sexo  
rename p207 gender
```

- Renombrando grupos de variables:

```
rename (age gender) (edad sexo)
```


El comando egen

Este comando permite crear nuevas variables a partir de estadísticos.

- Creamos la variable ingreso total cuyo contenido es la suma de diferentes variables:

| codperso | d529t | d536 | d540t | d543 | ing_extra | i530a | i524a1 | i538a1 | i541a | ingreso |
|----------|-------|------|-------|------|-----------|-------|--------|--------|-------|---------|
| 01 | . | 4063 | . | . | . | 7055 | . | . | . | 11118 |
| 02 | . | 5267 | . | 839 | . | 2786 | . | . | 572 | 9464 |
| 03 | . | . | . | . | . | . | . | . | . | 0 |

- Creamos una variable cuyo contenido sea el promedio del ingreso laboral por hogar:

| conglome | vivienda | hogar | ingreso | ingr_lab_hog |
|----------|----------|-------|---------|--------------|
| 005002 | 011 | 11 | 63271 | 21821.33 |
| 005002 | 011 | 11 | 2193 | 21821.33 |
| 005002 | 011 | 11 | 0 | 21821.33 |
| 005002 | 023 | 11 | 32175 | 32175 |
| 005002 | 049 | 11 | 23945 | 8118.25 |
| 005002 | 049 | 11 | 0 | 8118.25 |
| 005002 | 049 | 11 | 0 | 8118.25 |
| 005002 | 049 | 11 | 8528 | 8118.25 |

Cambiando tipo de almacenaje de variables

`destring`: convierte variables string a variables numéricas.

```
destring dpto, replace
```

| dpto | |
|------|----|
| 01 | 1 |
| 02 | 2 |
| 03 | 3 |
| 04 | 4 |
| 05 | 5 |
| 06 | 6 |
| 07 | 7 |
| 08 | 8 |
| 09 | 9 |
| 10 | 10 |



`tostring`: convierte variables numéricas a string.

```
tostring dpto, gen(str_dpto) force
```

| dpto | str_dpto |
|------|----------|
| 1 | 1 |
| 2 | 2 |
| 3 | 3 |
| 4 | 4 |
| 5 | 5 |
| 6 | 6 |
| 7 | 7 |
| 8 | 8 |
| 9 | 9 |
| 10 | 10 |

Cambiando tipo de almacenaje de variables

encode: pasa una variable string a una variable numérica etiquetando los valores string.

```
encode ubigeo, gen(id_ubigeo)
```

| ubigeo | id_ubigeo |
|--------|-----------|
| 010101 | 010101 |
| 021803 | 021803 |
| 030109 | 030109 |
| 040104 | 040104 |
| 050115 | 050115 |
| 060108 | 060108 |
| 070106 | 070106 |
| 080104 | 080104 |
| 090101 | 090101 |
| 100101 | 100101 |

decode: convierte variables numéricas con etiquetas en variables string.

```
decode dominio, gen(str_dominio)
```

| dominio | str_dominio |
|------------|--------------------|
| sierra ... | sierra norte |
| costa c... | costa centro |
| sierra ... | sierra sur |
| sierra ... | sierra sur |
| sierra ... | sierra centro |
| sierra ... | sierra norte |
| lima me... | lima metropolitana |
| sierra ... | sierra sur |
| sierra ... | sierra centro |
| sierra ... | sierra centro |

8. Manipulando Observaciones

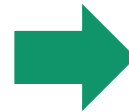
Reemplazar valores

- El comando “**replace**” reemplaza los valores de una variable ya creada.

Ingreso mensual

```
replace ingreso=ingreso/12
```

| conglome | vivienda | hogar | codperso | ingreso |
|----------|----------|-------|----------|---------|
| 005002 | 011 | 11 | 01 | 2193 |
| 005002 | 011 | 11 | 03 | 63271 |
| 005002 | 011 | 11 | 02 | 0 |
| 005002 | 023 | 11 | 01 | 32175 |
| 005002 | 049 | 11 | 01 | 0 |
| 005002 | 049 | 11 | 04 | 23945 |
| 005002 | 049 | 11 | 03 | 8528 |
| 005002 | 049 | 11 | 02 | 0 |



| conglome | vivienda | hogar | codperso | ingreso |
|----------|----------|-------|----------|----------|
| 005002 | 011 | 11 | 01 | 182.75 |
| 005002 | 011 | 11 | 03 | 5272.583 |
| 005002 | 011 | 11 | 02 | 0 |
| 005002 | 023 | 11 | 01 | 2681.25 |
| 005002 | 049 | 11 | 01 | 0 |
| 005002 | 049 | 11 | 04 | 1995.417 |
| 005002 | 049 | 11 | 03 | 710.6667 |
| 005002 | 049 | 11 | 02 | 0 |

Recodificar variables

El comando “**recode**” permite recodificar valores de una variable categórica y crear una nueva variable.

Primer paso: verificar la etiqueta de la variable.

```
. label list dominio
dominio:
      1 costa norte
      2 costa centro
      3 costa sur
      4 sierra norte
      5 sierra centro
      6 sierra sur
      7 selva
      8 lima metropolitana
```

Segundo paso: recodificar y crear una nueva variable

```
recode dominio (1/3=1 "Costa") (4/6=2 "Sierra") (7=3 "Selva") (8=4 "Lima Metropolitana"), gen(zona)
```

9. Borrar y Mantener Variables

El comando keep y drop

El comando “**keep**” mantiene las variables de interés.

```
* El comando keep
keep pais conglome vivienda hogar codperso dpto area zona pea p204 p205 p206 ///
edad sq_edad jefe mujer civil educ indigena pea peao informal ing_* ingreso ///
lnwage fac500a
```

| dpto | pais | ing_ocu_pri | ing_ocu_sec | ingreso | lnwage |
|------|------|-------------|-------------|----------|----------|
| 1 | Perú | 49035 | 0 | 5272.583 | 8.570275 |
| 1 | Perú | 0 | 0 | 0 | . |
| 1 | Perú | 2193 | 0 | 182.75 | 5.208119 |
| 1 | Perú | 17714 | 14461 | 2681.25 | 7.894038 |
| 1 | Perú | 0 | 0 | 0 | . |
| 1 | Perú | 0 | 0 | 0 | . |
| 1 | Perú | 8528 | 0 | 710.6667 | 6.566204 |
| 1 | Perú | 23335 | 0 | 1995.417 | 7.598608 |
| 1 | Perú | 0 | 0 | 0 | . |
| 1 | Perú | 0 | 0 | 0 | . |

El comando “**drop**” elimina las variables de interés.

```
* El comando drop
drop ing_*
```

| pais | ingreso | lnwage |
|------|----------|----------|
| Perú | 5272.583 | 8.570275 |
| Perú | 0 | . |
| Perú | 182.75 | 5.208119 |
| Perú | 2681.25 | 7.894038 |
| Perú | 0 | . |
| Perú | 0 | . |
| Perú | 710.6667 | 6.566204 |
| Perú | 1995.417 | 7.598608 |
| Perú | 0 | . |
| Perú | 0 | . |

10. Ordenar Observaciones y Variables

El comando sort y gsort

El comando “**sort**” ordena ascendentemente

```
sort ingreso
```

| conglome | vivienda | hogar | codperso | ingreso |
|----------|----------|-------|----------|----------|
| 018176 | 093 | 11 | 01 | 22771.17 |
| 015603 | 102 | 11 | 01 | 22778.58 |
| 017893 | 078 | 11 | 01 | 23715.42 |
| 007220 | 091 | 11 | 02 | 23967.92 |
| 018211 | 004 | 11 | 01 | 24961.08 |
| 005406 | 064 | 11 | 02 | 25449.5 |
| 019006 | 052 | 11 | 03 | 26155.33 |
| 008914 | 061 | 11 | 01 | 28198.33 |
| 008237 | 016 | 11 | 02 | 53323 |
| 005759 | 046 | 11 | 02 | 74092.16 |

El comando “**gsort**” ordena descendientemente

```
gsort -ingreso
```

| conglome | vivienda | hogar | codperso | ingreso |
|----------|----------|-------|----------|----------|
| 005759 | 046 | 11 | 02 | 74092.16 |
| 008237 | 016 | 11 | 02 | 53323 |
| 008914 | 061 | 11 | 01 | 28198.33 |
| 019006 | 052 | 11 | 03 | 26155.33 |
| 005406 | 064 | 11 | 02 | 25449.5 |
| 018211 | 004 | 11 | 01 | 24961.08 |
| 007220 | 091 | 11 | 02 | 23967.92 |
| 017893 | 078 | 11 | 01 | 23715.42 |
| 015603 | 102 | 11 | 01 | 22778.58 |
| 018176 | 093 | 11 | 01 | 22771.17 |

Subíndices de observaciones

- El subíndice `_n`: permite crear un correlativo desde la observación “1” hasta la observación “n”.
- El subíndice `_N`: permite crear una variable con el total de observaciones de la base de datos.



```
* El subíndice _n  
gen orden_obs= _n
```



```
* El subíndice _N  
gen total_obs=_N
```

| conglome | vivienda | hogar | codperso | orden_obs | total_obs |
|----------|----------|-------|----------|-----------|-----------|
| 005002 | 011 | 11 | 01 | 1 | 91315 |
| 005002 | 011 | 11 | 02 | 2 | 91315 |
| 005002 | 011 | 11 | 03 | 3 | 91315 |
| 005002 | 023 | 11 | 01 | 4 | 91315 |
| 005002 | 049 | 11 | 01 | 5 | 91315 |
| 005002 | 049 | 11 | 02 | 6 | 91315 |
| 005002 | 049 | 11 | 03 | 7 | 91315 |
| 005002 | 049 | 11 | 04 | 8 | 91315 |
| 005002 | 073 | 11 | 01 | 9 | 91315 |
| 005002 | 073 | 11 | 02 | 10 | 91315 |

El comando order

Es siempre recomendable ordenar las variables de nuestra base, tal que al inicio estén los identificadores de la unidad de observación y luego sus atributos.

```
order pais dpto zona area, before(conglome)
```

| pais | dpto | zona | area | conglome |
|------|------|--------|------|----------|
| Perú | 1 | Sierra | 1 | 005002 |
| Perú | 1 | Sierra | 1 | 005002 |
| Perú | 1 | Sierra | 1 | 005002 |
| Perú | 1 | Sierra | 1 | 005002 |
| Perú | 1 | Sierra | 1 | 005002 |
| Perú | 1 | Sierra | 1 | 005002 |
| Perú | 1 | Sierra | 1 | 005002 |
| Perú | 1 | Sierra | 1 | 005002 |
| Perú | 1 | Sierra | 1 | 005002 |
| Perú | 1 | Sierra | 1 | 005002 |

```
order ingreso lnwage fac500a, last
```

| ingreso | lnwage | fac500a | |
|----------|----------|----------|--|
| 182.75 | 5.208119 | 51.73801 | |
| 0 | . | 82.49285 | |
| 5272.583 | 8.570275 | 42.14626 | |
| 2681.25 | 7.894038 | 37.88447 | |
| 0 | . | 48.53986 | |
| 0 | . | 48.87992 | |
| 710.6667 | 6.566204 | 59.9916 | |
| 1995.417 | 7.598608 | 42.14626 | |
| 0 | . | 37.88447 | |
| 0 | . | 46.15303 | |

11. Etiquetas

El comando label variable

El comando “**label var**” permite asignar etiquetas a las variables.

```
label var pais "País"  
label var dpto "Departamento"  
label var zona "Zona Geográfica"  
label var area "Área Geográfica"  
label var mujer "Mujer"  
label var edad "Edad (años)"  
label var sq_edad "Edad al cuadrado"  
label var jefe "Jefe de hogar"  
label var civil "Estado Civil"  
label var educ "Nivel educativo"  
label var indigena "Indígena"
```

| variable name |
|---------------|
| pais |
| dpto |
| zona |
| area |
| mujer |
| edad |
| jefe |
| civil |
| educ |
| indigena |



| variable label |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| RECODE of dominio (dominio geográfico) RECODE of estrato (estrato geográfico) RECODE of sexo (sexo) ¿qué edad tiene en años cumplidos? RECODE of p203 (relación de parentesco con el jefe(a) del hogar) RECODE of p209 (¿cuál es su estado civil o conyugal?) RECODE of p301a (¿cuál es el último año o grado de estudios y nivel que aprobó?) RECODE of p558c (por sus antepasados y de acuerdo a sus costumbres, ¿ud. se cons |



| variable label |
|----------------------------------------------------------------------------------------------------------------------------------------------------|
| País Departamento Zona Geográfica Área Geográfica Mujer Edad (años) Jefe de hogar Estado Civil Nivel educativo Indígena |

El comando label define

El comando “**label define**” permite crear el nombre de la etiqueta con sus respectivos valores.

| dpto | area |
|------|------|
| 1 | 1 |
| 1 | 0 |
| 2 | 1 |
| 2 | 0 |
| 3 | 1 |
| 3 | 0 |
| 4 | 1 |
| 4 | 0 |
| 5 | 1 |
| 5 | 0 |
| 6 | 1 |
| 6 | 0 |
| 7 | 1 |
| 8 | 1 |
| 8 | 0 |

```
label define lab_area 1 "Urbano" 0 "Rural"
```

```
label define lab_dpto 1 "Amazonas" 2 "Ancash" 3 "Apurímac" 4 "Arequipa" ///  
5 "Ayacucho" 6 "Cajamarca" 7 "Callao" 8 "Cusco" 9 "Huancavelica" 10 "Huánuco" ///  
11 "Ica" 12 "Junín" 13 "La Libertad" 14 "Lambayeque" 15 "Lima" 16 "Loreto" ///  
17 "Madre de Dios" 18 "Moquegua" 19 "Pasco" 20 "Piura" 21 "Puno" 22 "San Martín" ///  
23 "Tacna" 24 "Tumbes" 25 "Ucayali"
```

El comando label values

El comando “`label values`” asocia la etiqueta creada en `label define` con una variable.

| dpto | area |
|------|------|
| 1 | 1 |
| 1 | 0 |
| 2 | 1 |
| 2 | 0 |
| 3 | 1 |
| 3 | 0 |
| 4 | 1 |
| 4 | 0 |
| 5 | 1 |
| 5 | 0 |
| 6 | 1 |
| 6 | 0 |
| 7 | 1 |
| 8 | 1 |
| 8 | 0 |



```
label values area lab_area  
label values dpto lab_dpto
```



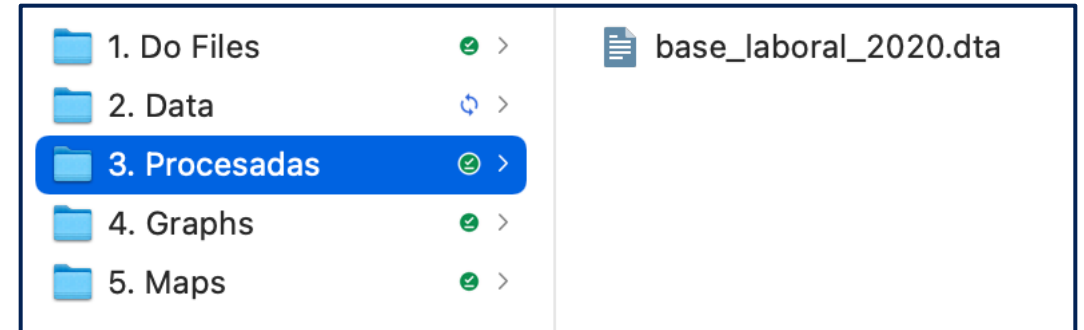
| dpto | area |
|-----------|--------|
| Amazonas | Urbano |
| Amazonas | Rural |
| Ancash | Urbano |
| Ancash | Rural |
| Apurímac | Urbano |
| Apurímac | Rural |
| Arequipa | Urbano |
| Arequipa | Rural |
| Ayacucho | Urbano |
| Ayacucho | Rural |
| Cajamarca | Urbano |
| Cajamarca | Rural |
| Callao | Urbano |
| Cusco | Urbano |
| Cusco | Rural |

12. Guardar y Exportar

El comando save

Una vez que hemos hecho los cambios deseados, podemos grabar la base de datos con el comando “**save**”.

```
* El comando save
*-----
save "base_laboral_2020.dta", replace
save "$works/base_laboral_2020.dta", replace
```



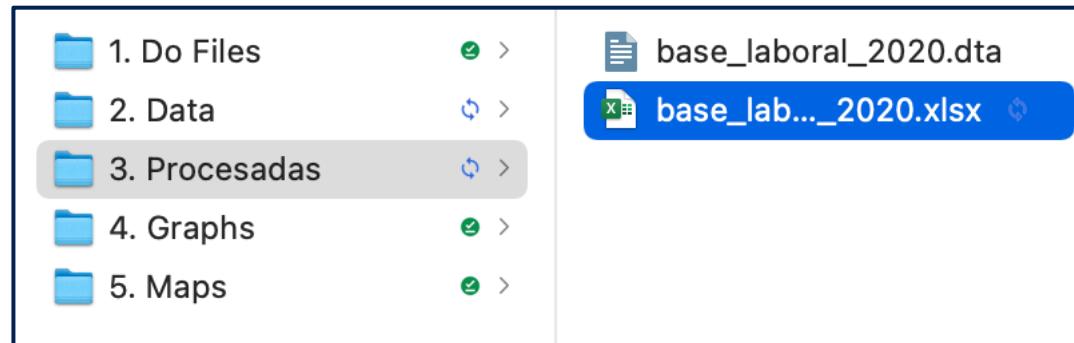
Exportar archivos

El comando “`export excel`” es especialmente diseñado para exportar la base en la memoria a Excel

* El comando `export excel`

*-----

```
export excel using "$works/base_laboral_2020.xlsx", firstrow(variables) replace
export excel using "$works/base_laboral_2020.xlsx" if dpto==15, firstrow(variables) sheet("lima")
export excel using "$works/base_laboral_2020.xlsx" in 1/7, firstrow(variables) sheet("selec_in")
```





PUCP