

Uso de paquetes estadísticos para el análisis de la pobreza y la desigualdad social en base a la encuesta permanente de hogares (EPH).

José Rodríguez de la Fuente.

Cita:

José Rodríguez de la Fuente (2018). *Uso de paquetes estadísticos para el análisis de la pobreza y la desigualdad social en base a la encuesta permanente de hogares (EPH)*. Material Didáctico Sistematizado.

Dirección estable: <https://www.aacademica.org/joserodriguez/70>

ARK: <https://n2t.net/ark:/13683/pq7B/wuz>



Esta obra está bajo una licencia de Creative Commons.
Para ver una copia de esta licencia, visite
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>.

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.

Uso de paquetes estadísticos.

Para el análisis de la pobreza y la desigualdad social en base a la encuesta permanente de hogares

JOSE RODRÍGUEZ DE LA FUENTE (IIGG – UBA – CONICET – INCASI)

Propuesta del taller

- Introducir características básicas y centrales de los paquetes SPSS y STATA
- Presentar la encuesta permanente de hogares (INDEC)
- Construir principales indicadores de mercado laboral
- Construir línea de pobreza e indigencia. Análisis a partir de grupos.
- Presentar indicadores de desigualdad social (brechas, gini, theil, etc.)

Breve historia de la EPH

- Primer relevamiento: Encuesta de Empleo y Desempleo (1963-1973). Primer EPH (1973 - GBA).
- ¿Qué mide?
 - Aspectos demográficos básicos
 - Características ocupacionales y de migraciones
 - Características habitacionales, educacionales y de ingresos.
- Continua: trimestral
- Encuesta de tipo panel: esquema de rotación (2-2-2)
- 3 Cuestionarios (vivienda, hogares, individuos) / 2 bases (individuos y hogares)
- Unidades de análisis: individuos, hogares, viviendas

Nombre del aglomerado**Jurisdicciones**

• Bahía Blanca - Cerri	Buenos Aires
• Comodoro Rivadavia - Rada Tilly	Chubut
• Concordia	Entre Ríos
• Corrientes	Corrientes
• Formosa	Formosa
• Gran Buenos Aires	Ciudad Autónoma de Buenos Aires y partidos del Gran Buenos Aires
• Gran Catamarca	Catamarca
• Gran Córdoba	Córdoba
• Gran La Plata	Buenos Aires
• Gran Mendoza	Mendoza
• Gran Paraná	Entre Ríos
• Gran Resistencia	Chaco
• Gran Rosario	Santa Fe
• Gran San Juan	San Juan
• Gran Santa Fe - Santo Tomé	Santa Fe
• Gran Tucumán - Tafi Viejo	Tucumán
• Jujuy - Palpalá	Jujuy
• La Rioja	La Rioja
• Mar del Plata - Batán	Buenos Aires
• Neuquén - Plottier	Neuquén
• Posadas	Misiones
• Rawson - Trelew	Chubut
• Río Cuarto	Córdoba
• Río Gallegos	Santa Cruz
• Salta	Salta
• San Luis - El Chorrillo	San Luis
• San Nicolás - Villa Constitución	Buenos Aires - Santa Fe
• Santa Rosa - Toay	La Pampa
• Santiago del Estero - La Banda	Santiago del Estero
• Ushuaia - Río Grande	Tierra del Fuego, Antártida e Islas del Atlántico Sur
• Viedma - Carmen de Patagones	Río Negro - Buenos Aires

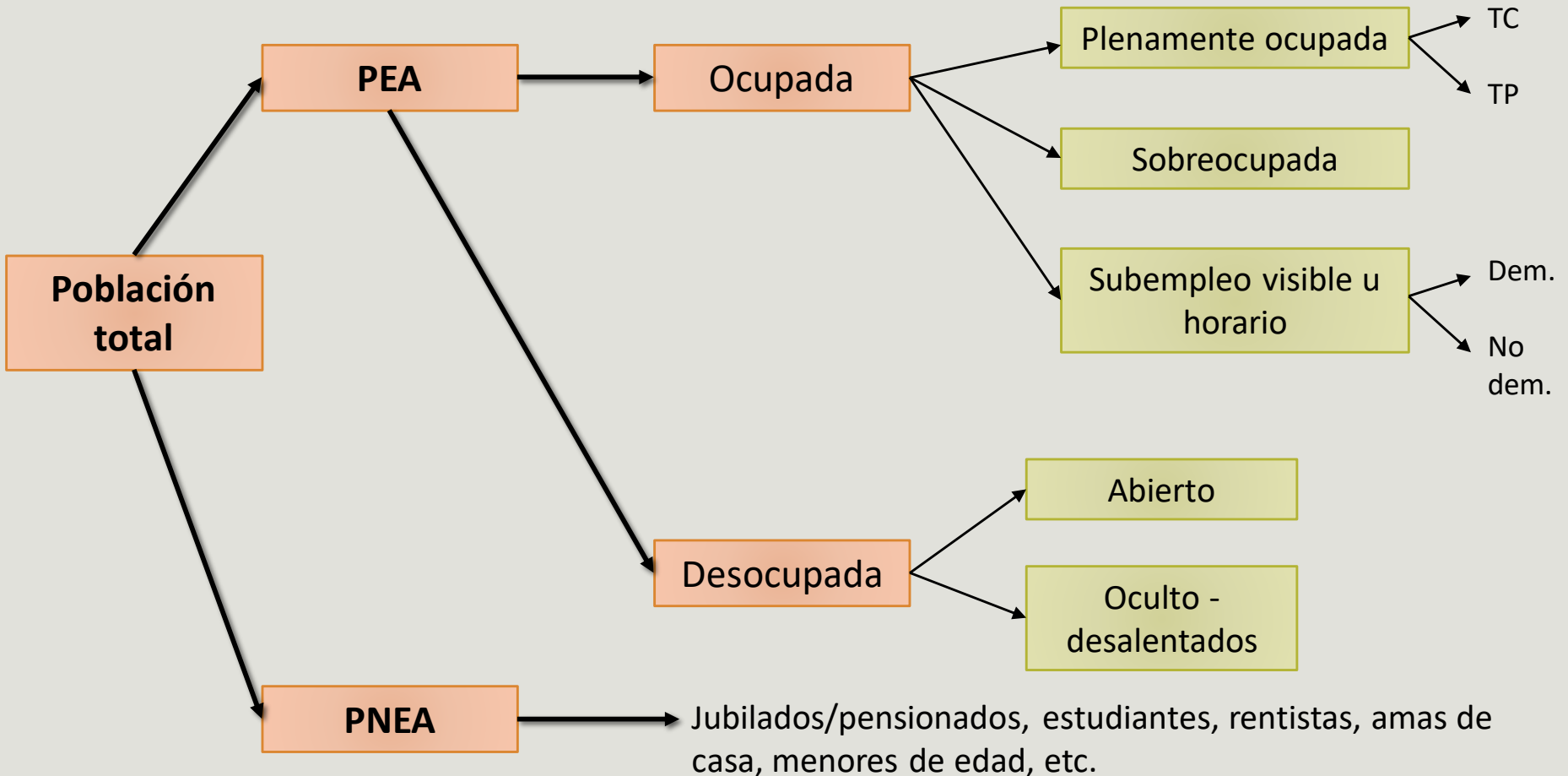
Aglomerados (2015)

- Incorporación de Mar del Plata: octubre 1995.
- Actualidad: 31 aglomerados urbanos y uno urbano-rural

Otras encuestas similares:

- EAHU (+ 2000 habitantes)
- ENAPROSS
- SIPA (no es encuesta)
- Encuestas provinciales
- PISAC (Inter-universitario)

Categorías de la fuerza de trabajo captadas por la EPH



Bases INDEC

- Hasta 2015 los datos se encontraban en formato SPSS, STATA y DBF
- Desde 2016 se publica sólo en formato xls y csv ¿Problemas?
- Necesidad de importar los archivos y colocar las etiquetas a las variables y categorías
- Formato rápido de consulta: REDATAM

Encuesta Permanente de Hogares (EPH)

Microdatos y documentos 2016 y 2017 ▼

Cuadros regulares - EPH Continua 2003-2006 y 2016-2017 ▼

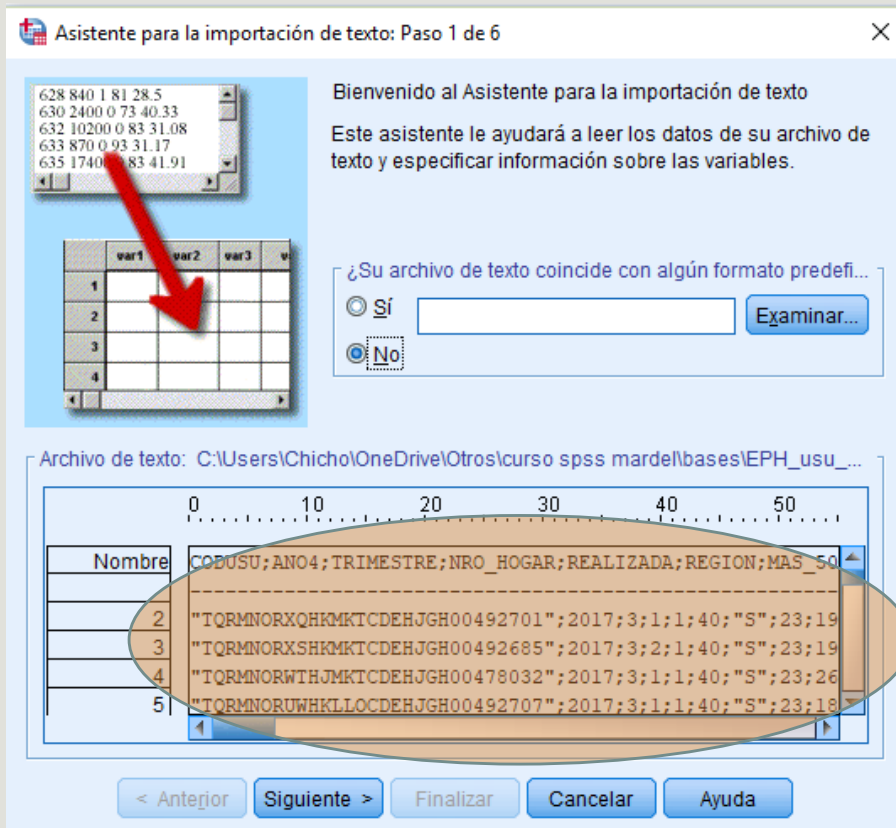
Bases REDATAM 2003-2015 ▼

Microdatos y documentos 2003-2015 ▼

Información histórica (2003-2006) ▼

Utilizando la EPH en SPSS (I)

- Abrir base -> Pasar por el asistente de importación



Base sin formato

Posteriormente la base puede guardarse en varios formatos. Recomendable usar el formato .SAV

Editor de datos (modalidad vista de datos)

Visible: 88 de 88 variables

	CODUSU	ANO4	TRIMEST RE	NRO_HO GAR	REALI	ESCDU	MAS_500	AGLOME RADO	PONDER A	IV1	IV1_ESP	IV2	IV3
1	TQRMNORXQHKMKTCDEHJGH00492701	2017	3	1	1	40 S		23	191	2		3	1
2	TQRMNORXSHKMKTCDEHJGH00492685	2017	3	2	1	40 S		23	191	2		2	1
3	TQRMNORWTHJMKTCDEHJGH00478032	2017	3	1	1	40 S		23	264	1		3	1
4	TQRMNORUWHKLLCDEHJGH00492707	2017	3	1	1	40 S		23	185	1		4	1
5	TQRMNORWUHMLUCDEHJGH00597507	2017	3	1	1	40 S		23	240	1		4	1
6	TQRMNOPWPHKMLNCDEHJGH00492714	2017	3	1	1	40 S		23	213	1		3	2
7	TQRMNORXGHNKMLNCDEHJGH00492724	2017	3	2	1	40 S		23	213	1		5	2
	TQRMNORXYHKNLNCDEHJGH00492715	2017	3	1	1	40 S		23	213	1		6	1
	TQRMNOQVWHJMKUCDEHJGH00478039	2017	3	1	1	40 S		23	249	1		5	1
	TQRMNORTTHKMLPCDEHJGH00492735	2017	3	1	1	40 S		23	219	1		2	1
	TQRMNOPWWHMNLQCDEHJGH00597508	2017	3	1	1	40 S		23	226	1		5	2
	TQRMNOPWYHMLNCDEHJGH00597509	2017	3	1	1	40 S		23	226	1		2	2
	TQRMNOSRQHKMLRCDEHJGH00492728	2017	3	1	1	40 S		23	243	2		3	1
	TQRMNOSRTHKMLRCDEHJGH00492669	2017	3	2	1	40 S		23	243	2		3	1
	TQRMNOPYUHMMLRCDEHJGH00597510	2017	3	1	1	40 S		23	241	1		5	1
	TQRMNOPXSHMLQCDEHJGH00478060	2017	3	1	1	40 S		23	242	1		4	2
17	TQRMNORXVHKMLSCDEHJGH00492565	2017	3	1	1	40 S		23	224	1		2	2
18	TQRMNORXGHNKMLSCDEHJGH00597481	2017	3	1	1	40 S		23	213	1		4	2
19	TQRMNOSYUHMMLKPCDEHJGH00597511	2017	3	1	1	40 S		23	211	2		4	1
20	TQRMNORTRHMLQCDEHJGH00597512	2017	3	1	1	40 S		23	233	1		5	1
21	TQRMNORQSHMLQCDEHJGH00597513	2017	3	1	1	40 S		23	226	1		6	2
22	TQRMNOQVQHKOLQCDEHJGH00492691	2017	3	1	1	40 S		23	198	1		1	2

Registros

Valores

Vista de datos | Vista de variables

IBM SPSS Statistics Processor está listo | Unicode: ON

10:34 p.m. 26/3/2018

Editor de datos (modalidad vista de variables)

*Sin titulo2 [ConjuntoDatos1] - IBM SPSS Statistics Editor de datos

Archivo Editar Ver Datos Transformar Analizar Gráficos Utilidades Ampliaciones Ventana Ayuda

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	CODUSU	Cadena	29	0		Ninguno	Ninguno	29	Izquierda	Nominal	Entrada
2	ANO	Numérico	4	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
3	TRIMESTRE	Numérico	1	0		Ninguno	Ninguno	8	Derecha	Nominal	Entrada
4	NRO_HOGAR	Numérico	2	0		Ninguno	Ninguno	8	Derecha	Nominal	Entrada
5	REALIZADA	Numérico	1	0		Ninguno	Ninguno	8	Derecha	Nominal	Entrada
6	REGION	Numérico	2	0		Ninguno	Ninguno	8	Derecha	Nominal	Entrada
7	MAS_500	Cadena	1	0		Ninguno	Ninguno	9	Izquierda	Nominal	Entrada
8	AGLOMER...	Numérico	2	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
9	PONDERA	Numérico	4	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
10	IV1	Numérico	1	0		Ninguno	Ninguno	8	Derecha	Nominal	Entrada
11	IV1_ESP	Numérico	1	0		Ninguno	Ninguno	8	Derecha	Nominal	Entrada
12	IV2	Numérico	2	0		Ninguno	Ninguno	8	Derecha	Nominal	Entrada
13	IV3	Numérico	1	0		Ninguno	Ninguno	8	Derecha	Nominal	Entrada
14	IV3_ESP	Numérico	1	0		Ninguno	Ninguno	8	Derecha	Nominal	Entrada
15	IV4	Numérico	1	0		Ninguno	Ninguno	8	Derecha	Nominal	Entrada
16	IV5	Numérico	1	0		Ninguno	Ninguno	8	Derecha	Nominal	Entrada
17	IV6	Numérico	1	0		Ninguno	Ninguno	8	Derecha	Nominal	Entrada
18	IV7	Numérico	1	0		Ninguno	Ninguno	8	Derecha	Nominal	Entrada
19	IV7_ESP	Numérico	1	0		Ninguno	Ninguno	8	Derecha	Nominal	Entrada
20	IV8	Numérico	1	0		Ninguno	Ninguno	8	Derecha	Nominal	Entrada
21	IV9	Numérico	1	0		Ninguno	Ninguno	8	Derecha	Nominal	Entrada
22	IV10	Numérico	1	0		Ninguno	Ninguno	8	Derecha	Nominal	Entrada
23	IV11	Numérico	1	0		Ninguno	Ninguno	8	Derecha	Nominal	Entrada
24	IV12_1	Numérico	1	0		Ninguno	Ninguno	8	Derecha	Nominal	Entrada
25	IV12_2	Numérico	1	0		Ninguno	Ninguno	8	Derecha	Nominal	Entrada

Valores y etiquetas

IBM SPSS Statistics Processor está listo

Unicode:ON

11:02 p.m.
26/3/2018

Nombre variable

Etiqueta variable

Escala de medición de la variable

Visor de resultados

Estadísticos

DECIFR

N	Válido	18806
	Perdidos	0
Moda		12
Rango		12
Mínimo		0
Máximo		12

DECIFR

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido 0	91	,5	,5	,5
1	1859	9,9	9,9	10,4
2	1895	10,1	10,1	20,4
3	1441	7,7	7,7	28,1
4	1633	8,7	8,7	36,8
5	1359	7,2	7,2	44,0
6	1566	8,3	8,3	52,3
7	1330	7,1	7,1	59,4
8	1484	7,9	7,9	67,3
9	1254	6,7	6,7	74,0
10	1239	6,6	6,6	80,6
12	3655	19,4	19,4	100,0
Total	18806	100,0	100,0	

Puedo recortar y pegar en Word o Excel los resultados para formatearlos y utilizarlos

IBM SPSS Statistics Processor está listo Unicode:ON

11:07 p.m.
26/3/2018

Editor de sintaxis

*Sintaxis1 - IBM SPSS Statistics Editor de sintaxis

Archivo Editar Ver Datos Transformar Analizar Gráficos Utilidades Ejecutar Herramientas Ampliaciones Ventana Ayuda

Activo: ConjuntoDatos1

```
1  
2 DATASET ACTIVATE ConjuntoDatos1.  
3 FREQUENCIES VARIABLES=DECCFR  
4 /ORDER=ANALYSIS.  
5
```

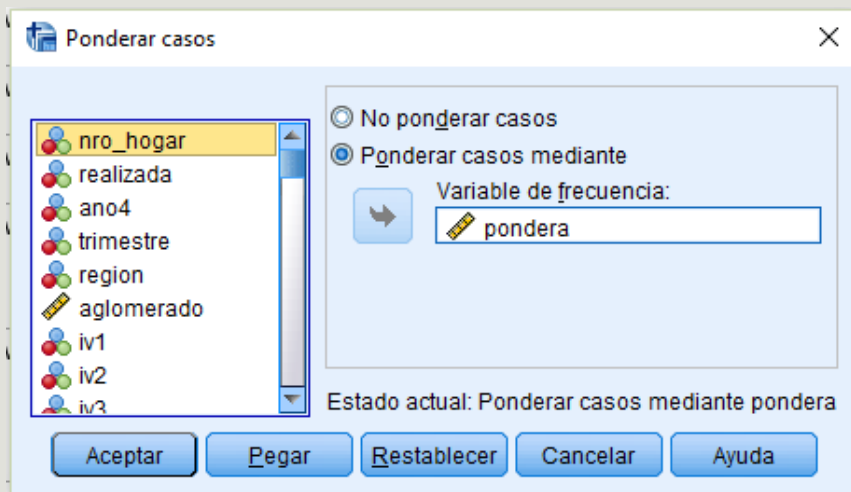
IBM SPSS Statistics Processor está listo Unicode:ON In 4 Col 18 NUM

11:10 p.m.
26/3/2018

- Funciona como procesador de texto
- Solo funciona cuando se da la orden de ejecución
- Se puede ejecutar la totalidad de la sintaxis o sólo una parte
- Permite hacer anotaciones con el símbolo *
- Es útil para el proceso investigativo: la sintaxis guarda los procesamientos que luego pueden ser leídos y corroborados por otros investigadores

Utilizando la EPH en SPSS (II)

- Antes que nada, para utilizar correctamente la base es necesario ponderarla
- La variable “pondera”, funciona como ponderador y como factor de expansión.
- Se agregaron nuevos ponderadores a partir de 2016.



Comando RECODE (Recodificar)

- Permite la agrupación de categorías de las variables
- Permite hacer la recategorización en la misma variable o en una distinta
- Probemos con la variable estado (condición de actividad) de la base individual
- Transformar -> Recodificar en distinta variable

Recodificar en distintas variables

Variable numérica -> Variable de salida:
estado -> pea

Variable de salida
Nombre: pea
Etiqueta: Población económicamen
[Cambiar]

Valores antiguos y nuevos...
Si... (condición de selección de caso opcional)

Aceptar Pegar Restablecer Cancelar Ayuda

Recodificar en variables diferentes: valores antiguo y nuevo

Valor antiguo:
 Valor:
[]
 Perdido del sistema
 Perdido por el sistema o el usuario

Rango:
[] hasta []
 Rango, LOWEST hasta el valor:
[]
 Rango, valor hasta HIGHEST:
[]
 Todos los demás valores

Valor nuevo:
 Valor: []
 Perdido del sistema
 Copiar valores antiguos

Antiguo -> Nuevo:
1 -> 1
2 -> 1
3 -> 2
4 -> 2

Añadir
Cambiar
Eliminar

Las variables de salida son series Anchura: 8
 Convertir series numéricas a números ('5'->5)

Continuar Cancelar Ayuda

Etiquetar categorías de variables

167	deccfr	Cadena	2	0	Grupo decilic...	Ninguno	Ninguno	8	Izquierda	Nominal	Entrada
168	ideccfr	Cadena	2	0	Grupo decilic...	Ninguno					
169	rdeccfr	Cadena	2	0	Grupo decilic...	Ninguno					
170	gdeccfr	Cadena	2	0	Grupo decilic...	Ninguno					
171	pdeccfr	Cadena	2	0	Grupo decilic...	Ninguno					
172	adeccfr	Cadena	2	0	Grupo decilic...	Ninguno					
173	idimpp	Cadena	5	0	Identificador ...	{00000, No i...					
174	pea	Numérico	8	2	Población ec...	Ninguno					
175											
176											
177											
178											
179											
180											
181											
182											

Etiquetas de valor

Etiquetas de valor

Valor:

Etiqueta:

1,00 = "PEA"
2,00 = "PNEA"

Añadir
Cambiar
Eliminar

Ortografía...

Aceptar Cancelar Ayuda

Forma prolija (en sintaxis):

VALUE LABELS pea

1 "PEA"

2 "PNEA".



		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	1,00 PEA	11920216	44,4	44,5	44,5
	2,00 PNEA	14893458	55,5	55,5	100,0
	Total	26813674	99,9	100,0	
Perdidos	Sistema	34467	,1		
Total		26848141	100,0		

Comando COMPUTE (calcular)

- Permite construir una nueva variable a partir de la combinación de otras
- Permite hacer la recategorización en la misma variable o en una distinta
- Puedo construir variables directamente usando operadores aritméticos y lógicos o a partir de condiciones
- Probemos con intentar construir la variable Intensidad laboral
- Transformar -> Calcular variable

The image displays two screenshots of the SPSS 'Calcular variable' (Calculate Variable) dialog box, illustrating the process of creating a new variable based on existing ones and applying a condition.

Left Screenshot: 'Calcular variable'

- Variable objetivo:** totalhoras
- Expresión numérica:** pp3e_tot + pp3f_tot
- Lista de variables:** pp02i, pp03c, pp03d, pp3e_tot, pp3f_tot, pp03g, pp03h, pp03i, pp03j, intensi, pp04a, pp04b_caes, pp04b1, pp04b2, pp04b3_mes, pp04b3_ano, pp04b3_dia, pp04c, nn04c99
- Grupo de funciones:** Todo, Aritméticas, CDF y CDF no centrada, Conversión, Fecha/hora actual, Cálculo de fechas
- Funciones y variables especiales:** (Empty)
- Condición de selección de caso (opcional):** Si...

Right Screenshot: 'Calcular variable: Si los casos'

- Radio buttons:** Incluir todos los casos; Incluir si el caso satisface la condición
- Condición:** estado = 1 & (totalhoras >= 35 & totalhoras <= 45)
- Lista de variables:** adecindr, v2_m, v3_m, v4_m, v5_m, v8_m, v9_m, v10_m, v11_m, v12_m, v18_m, v19_am, v21_m, t_vi, if, decifr, idecifr, rdecifr, gdecifr, pdecifr, adecifr, ipcf, decfcr, ideccfcr, rdecfcr, gdecfcr, pdecfcr, adecfcr, idimpp, totalhoras, intensidad
- Grupo de funciones:** Todo, Aritméticas, CDF y CDF no centrada, Conversión, Fecha/hora actual, Cálculo de fechas, Creación de fechas, Extracción de fechas
- Funciones y variables especiales:** (Empty)

Cantidad de horas trabajadas	Trabajar más horas			
	Si		No	
	Buscó empleo		Buscó empleo	
	Si	No	Si	No
<35	3	4	1	1
35 / 45	1	1	1	1
>45	2	2	2	2

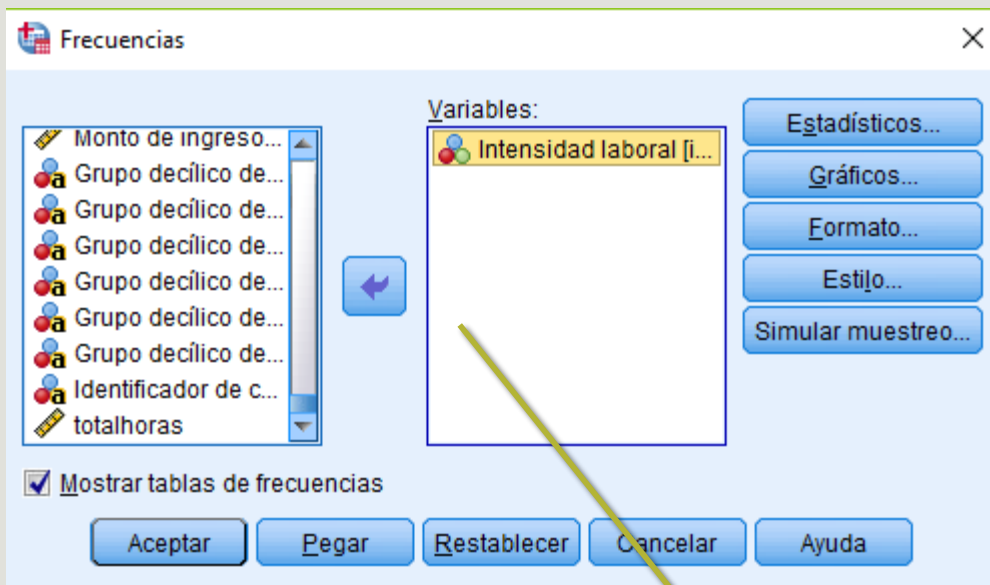
if estado = 1 & (totalhoras >= 35 & totalhoras <= 45) intensidad = 1.

if estado = 1 & totalhoras < 35 & pp03g = 2 intensidad = 1.

if estado = 1 & totalhoras > 45 intensidad = 2.

if estado = 1 & totalhoras < 35 & pp03g = 1 & pp03j = 1 intensidad = 3.

if estado = 1 & totalhoras < 35 & pp03g = 1 & pp03j = 2 intensidad = 4.

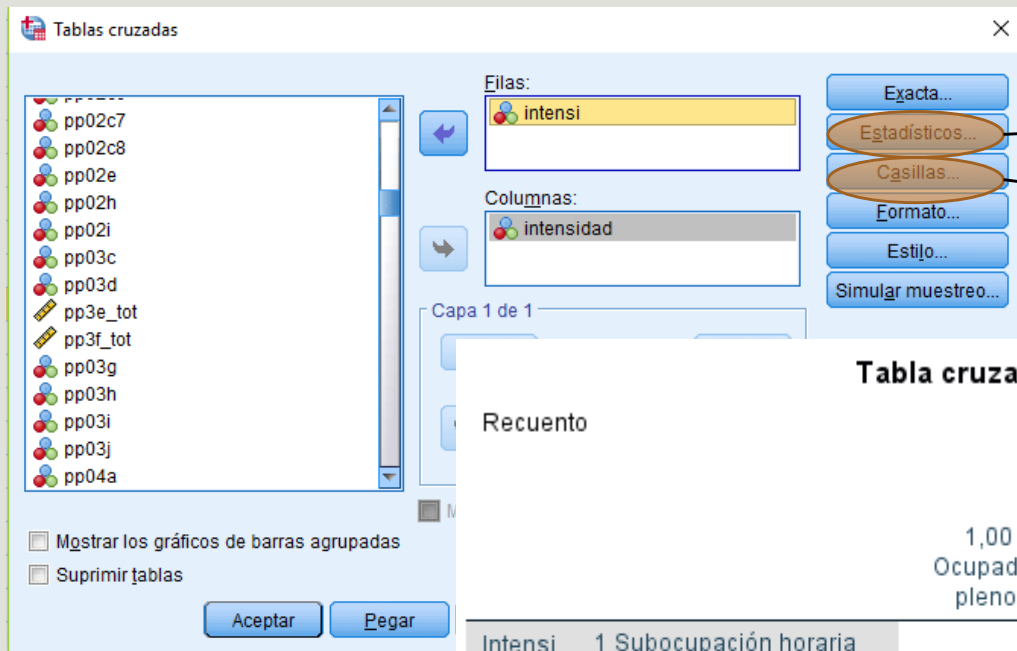


- Analizar -> Estadísticos descriptivos -> Frecuencias

Intensidad laboral

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	1,00 Ocupados plenos	6558786	24,4	58,9	58,9
	2,00 Sobreocupados	3446472	12,8	31,0	89,9
	3,00 Subocupados demandantes	765777	2,9	6,9	96,8
	4,00 Subocupados no demandantes	361310	1,3	3,2	100,0
	Total	11132345	41,5	100,0	
Perdidos	Sistema	15715796	58,5		
Total		26848141	100,0		

- Utilizando el comando “tablas cruzadas” podemos comparar la variable que creamos con la variable “intensi” que ya viene en la base EPH.
- Analizar → Estadísticos descriptivos → Tablas cruzadas



Estadísticos de asociación

Tipos de porcentajes

Tabla cruzada Intensi*Intensidad laboral

Recuento		Intensidad laboral				Total
		1,00 Ocupados plenos	2,00 Sobreocupad os	3,00 Subocupados demandantes	4,00 Subocupados no demandantes	
Intensi	1 Subocupación horaria Demandante	0	0	747308	0	747308
	2 Subocupación horaria No Demandante	0	0	0	329657	329657
	3 Ocupación plena	6344982	0	6207	18791	6369980
	4 Sobreocupación horaria	0	3385948	0	0	3385948
	5 Ocupado que no trabajó en la semana	213804	0	12262	12862	238928
	9 Ns./Nr.	0	60524	0	0	60524
Total		6558786	3446472	765777	361310	11132345

Calculo de principales tasas para el aglomerado Mar del Plata - Batán

Tasa de actividad: $(PEA / Población\ total) * 100$

Tasa de empleo: $(Ocupados / Población\ total) * 100$

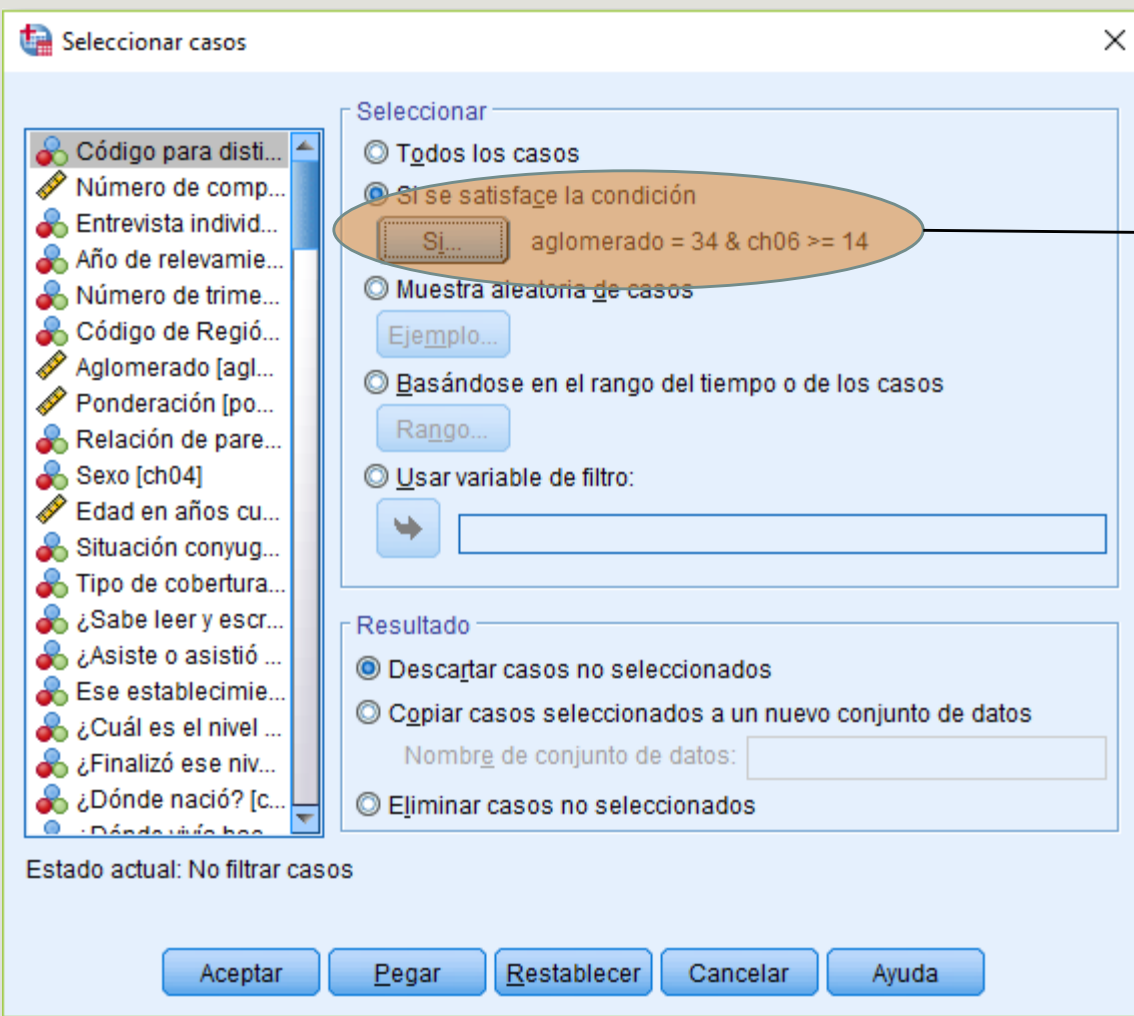
Tasa de desocupación: $(desocupados / PEA) * 100$

Tasa de subocupación: $(subocupados / PEA) * 100$

Tasa de sobreocupación: $(sobreocupados / PEA) * 100$

Las tasas pueden refinarse acotando a la población a la que será potencialmente activa (+ 14 años)

→ Necesidad de construir un filtro:
Datos → Seleccionar casos



Individuos que residen en el aglomerado Mar del Plata – Batán y son mayores o iguales a 14 años

Población Económicamente Activa

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	1,00 "PEA"	276361	55,0	55,0	55,0
	2,00 "PNEA"	226258	45,0	45,0	100,0
	Total	502619	100,0	100,0	

Tasa de actividad

Condición de actividad

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	0 Entrevista individual no realizada (no respuesta al cuestion)	3625	,7	,7	,7
	1 Ocupado	254029	50,5	50,5	51,3
	2 Desocupado	22332	4,4	4,4	55,7
	3 Inactivo	222633	44,3	44,3	100,0
	Total	502619	100,0	100,0	

Tasa de empleo

Condición de actividad

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	1 Ocupado	254029	91,9	91,9	91,9
	2 Desocupado	22332	8,1	8,1	100,0
	Total	276361	100,0	100,0	

Tasa de desocupación
(en el filtro sólo considero la PEA)

Intensi

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	0	22332	8,1	8,1	8,1
	1 Subocupación horaria Demandante	26467	9,6	9,6	17,7
	2 Subocupación horaria No Demandante	12166	4,4	4,4	22,1
	3 Ocupación plena	125276	45,3	45,3	67,4
	4 Sobreocupación horaria	80040	29,0	29,0	96,4
	5 Ocupado que no trabajó en la semana	10080	3,6	3,6	100,0
	Total	276361	100,0	100,0	

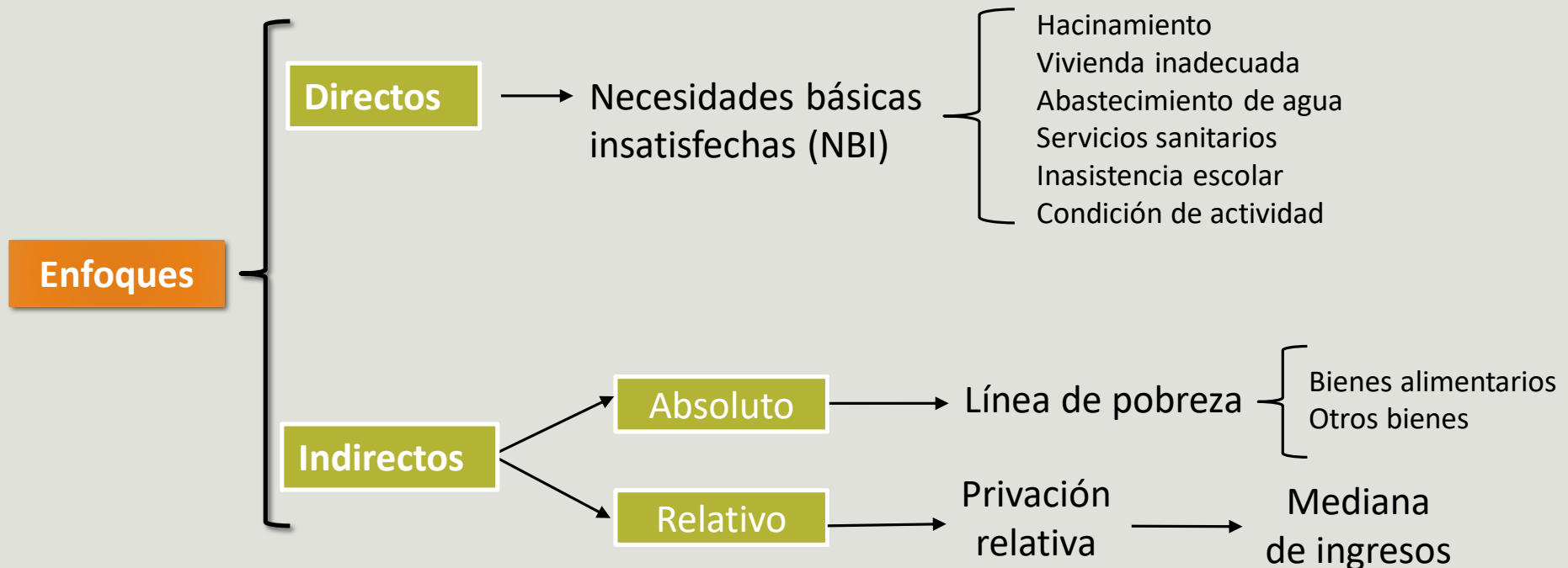
Tasa de subocupación

Tasa de sobreocupación

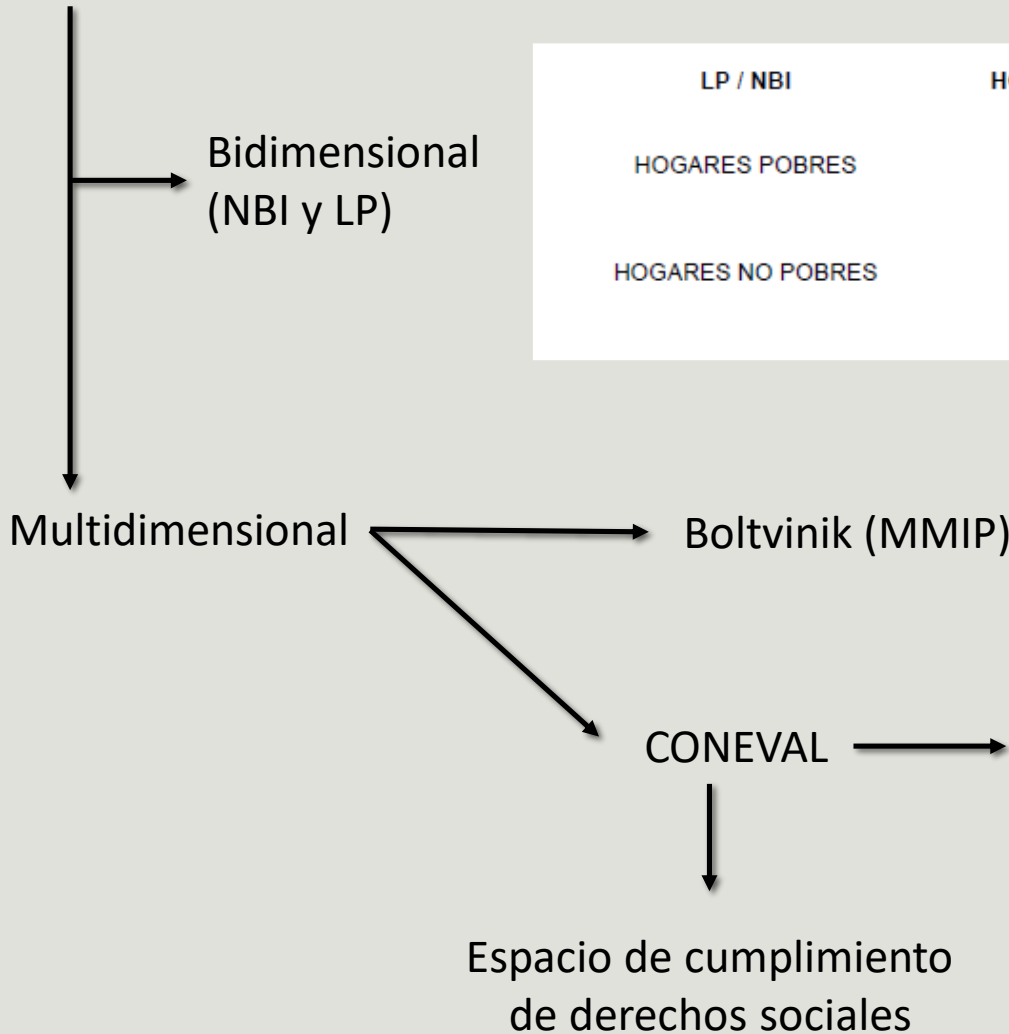
Medición de la pobreza e indigencia a partir SPSS

¿Qué es la pobreza?

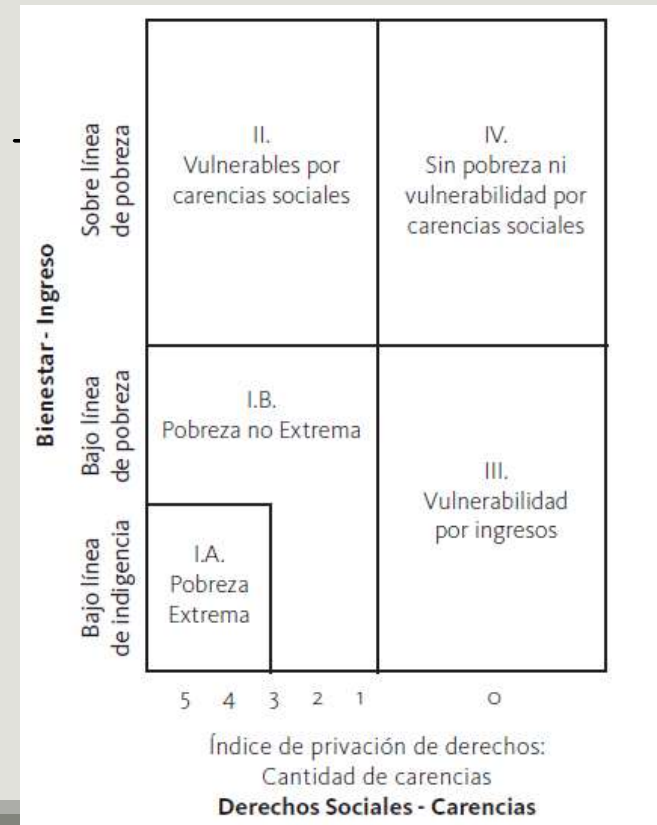
- Incapacidad de las personas de vivir una vida tolerable (PNUD, 1997)
- Dificultad para medir conceptos complejos como “nivel de vida” o “bienestar”



Enfoques combinados

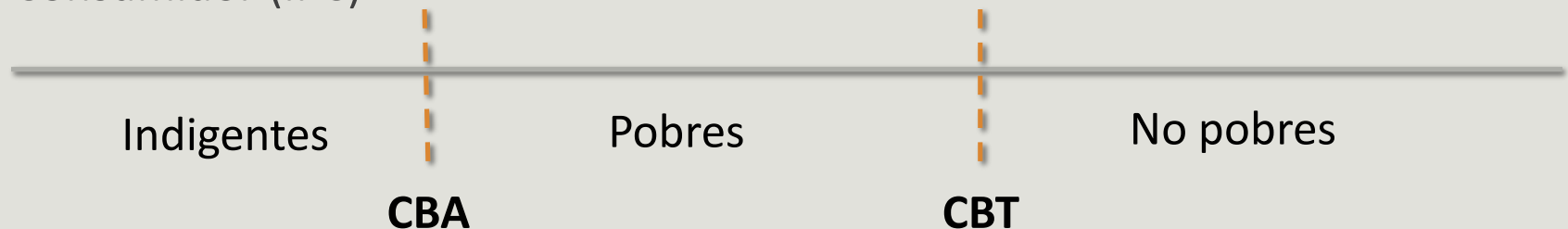


LP / NBI	HOGARES CON NBI	HOGARES SIN NBI
HOGARES POBRES	Pobreza Crónica (Total)	Pobreza Reciente (Pauperizados o Coyuntural)
HOGARES NO POBRES	Pobreza Inercial (Estructural)	Hogares en Condiciones de Integración Social



¿Cómo se calcula la pobreza en Argentina?

- Utilización de la medición absoluta: línea de pobreza y línea de indigencia
- La línea de indigencia se construye a partir del valor de la CANASTA BÁSICA ALIMENTARIA
- La línea de pobreza se construye a partir del valor de la CANASTA BÁSICA TOTAL, que incluye los bienes no alimentarios
- Los elementos que componen las canastas se basan en los hábitos de consumo relevados por las Encuestas de Gastos de Hogares (ENGHO).
- El valor se actualiza mensualmente a partir del Índice de Precios al Consumidor (IPC)



Cuadro 3. Canasta Básica Alimentaria.
Composición de la canasta para el adulto equivalente

Componente	Unidades	Productos que se incluyen
Pan	6.750 g	
Galletitas de agua	420 g	
Galletitas dulces	210 g	
Arroz	1.200 g	
Harina de trigo	1.080 g	
Otras harinas (maíz)	210 g	
Fideos	1.740 g	
Papa	6.510 g	
Batata	510 g	
Azúcar	1.230 g	
Dulces	330 g	Dulce de batata, mermelada, dulce de leche
Legumbres secas	240 g	Lentejas, arvejas
Hortalizas	5.730 g	Acelga, cebolla, lechuga, tomate perita, zanahoria, zapallo, tomate envasado
Frutas	4.950 g	Manzana, mandarina, naranja, banana, pera
Carnes	6.270 g	Asado, camaza común, espinazo, paleta, carne picada, nalga, pollo, carne de pescado
Menudencias	270 g	Higado
Fiambres	60 g	Paleta cocida, salame
Huevos	600 g	
Leche	9.270 g	
Queso	330 g	Queso crema, queso cuartirolo, queso de rallar
Yogur	570 g	
Manteca	60 g	
Aceite	1.200 g	
Bebidas no alcohólicas	3.450 cc	Gaseosas, jugos concentrados, soda
Bebidas alcohólicas	1.080 cc	Cerveza, vino
Sal fina	120 g	
Condimentos	120 g	Mayonesa, caldos concentrados
Vinagre	60 g	
Café	30 g	
Yerba	510 g	

Fuente: INDEC. Actualización de la Metodología oficial de cálculo de las líneas de pobreza. Argentina, 2004. Documento en discusión.

Serie Canasta Básica Alimentaria. Línea de Indigencia

Región	2017					
	Enero	Febrero	Marzo	Abril	Mayo	Junio
	Pesos					
Gran Buenos Aires	1.789,15	1.821,02	1.876,56	1.915,47	1.937,24	1.956,42
Cuyo	1.595,31	1.628,44	1.681,49	1.713,91	1.731,70	1.747,95
Noreste	1.590,44	1.620,35	1.672,78	1.709,08	1.729,40	1.743,77
Noroeste	1.541,73	1.570,68	1.623,37	1.661,21	1.681,95	1.698,68
Pampeana	1.777,09	1.810,32	1.863,58	1.901,16	1.922,00	1.940,78
Patagónica	1.848,50	1.882,42	1.939,76	1.975,67	1.994,94	2.012,13

¿Pero la CBA a qué unidad de medida responde?



Adulto equivalente: diferencias por sexo, edad y actividad de las personas en el consumo de energía y proteínas

La referencia es un varón adulto de entre 30 y 60 años de edad, con una actividad moderada



Sexo	Edad	Necesidad energética (kcal)	Unidades consumidoras
Ambos	6-9 meses	776	0,28
	9-12 meses	952	0,35
	1 año	1030	0,37
	2 años	1277	0,46
	3 años	1409	0,51
	4 años	1516	0,55
	5 años	1643	0,60
	6 años	1760	0,64
	7 años	1813	0,66
	8 años	1865	0,68
Varones	9 años	1910	0,69
	10 años	2192	0,79
	11 años	2255	0,82
	12 años	2347	0,85
	13 años	2472	0,90
	14 años	2650	0,96
	15 años	2760	1,00
	16 años	2826	1,03
Mujeres	17 años	2881	1,04
	10 años	1916	0,70
	11 años	1966	0,72
	12 años	2051	0,74
	13 años	2089	0,76
	14 años	2100	0,76
	15 años	2116	0,77
16 años	2111	0,77	
Varones	17 años	2124	0,77
	18-29 años	2826	1,02
	30-45 años	2758	1,00
	46-60 años	2750	1,00
	61-75 años	2288	0,83
Mujeres	Más de 75 años	2050	0,74
	18-29 años	2106	0,76
	30-45 años	2111	0,77
	46-60 años	2090	0,76
	61-75 años	1860	0,67
	Más de 75 años	1750	0,63

Hogar típico:

- Varón adulto (40 años): 1,00
- Mujer adulta (40 años): 0,77
- Niño (9 años): 0,69
- Niña (5 años): 0,60

Total: 3,06

¿Cuántos ingresos necesita ese hogar para no ser indigente?

CBA * Adulto equiv.

$$\$1901,16 * 3,06 = \$5817,5$$

Fuente: INDEC, "Actualización de la metodología oficial de cálculo de las Líneas de Pobreza" presentado en el Taller Regional de Medición de la Pobreza, Buenos Aires, noviembre de 2003.

¿Cómo se calcula la canasta básica total?

- Necesidad de tomar el valor de la CBA y ampliarlo incluyendo otros bienes no alimentarios (vestimenta, transporte, educación, salud, etc.)
- Utilización del “coeficiente de Engel” (CdE) definido como la relación entre los gastos alimentarios y los gastos totales observados a partir de la ENGHo 2004/2005.
- También el valor del CdE es regional y es actualizado mensualmente.
- $CBT = CBA * \text{inversa del CdE}$

Serie inversa del coeficiente de Engel

Región	2017					
	Enero	Febrero	Marzo	Abril	Mayo	Junio
Gran Buenos Aires	2,41	2,43	2,43	2,45	2,45	2,45
Cuyo	2,57	2,59	2,58	2,60	2,61	2,61
Noreste	2,26	2,28	2,27	2,29	2,29	2,30
Noroeste	2,24	2,26	2,25	2,27	2,27	2,28
Pampeana	2,41	2,43	2,43	2,44	2,45	2,45
Patagónica	2,74	2,76	2,75	2,78	2,78	2,79

Siguiendo con nuestro ejemplo:

¿Cuánto ingreso necesitaría el hogar típico para no ser considerado pobre?

CBA * inversa CdE * Adulto equiv

$$\text{\$1901,16} * 2,44 * 3,06 = \text{\$14.184, 8}$$

¿Cómo calculamos la pobreza a partir de SPSS?

- Aprendizaje de comandos: Agregar, fusionar, do if.
- Utilización de la base individuos (calculo de unidades de referencia) y de la base hogares (valor del Ingreso Total Familiar)
- Lo que se contrasta es el Ingreso Total Familiar (ITF) de los hogares con el valor de la línea de pobreza e indigencia para cada hogar.

1) Armar variable en función

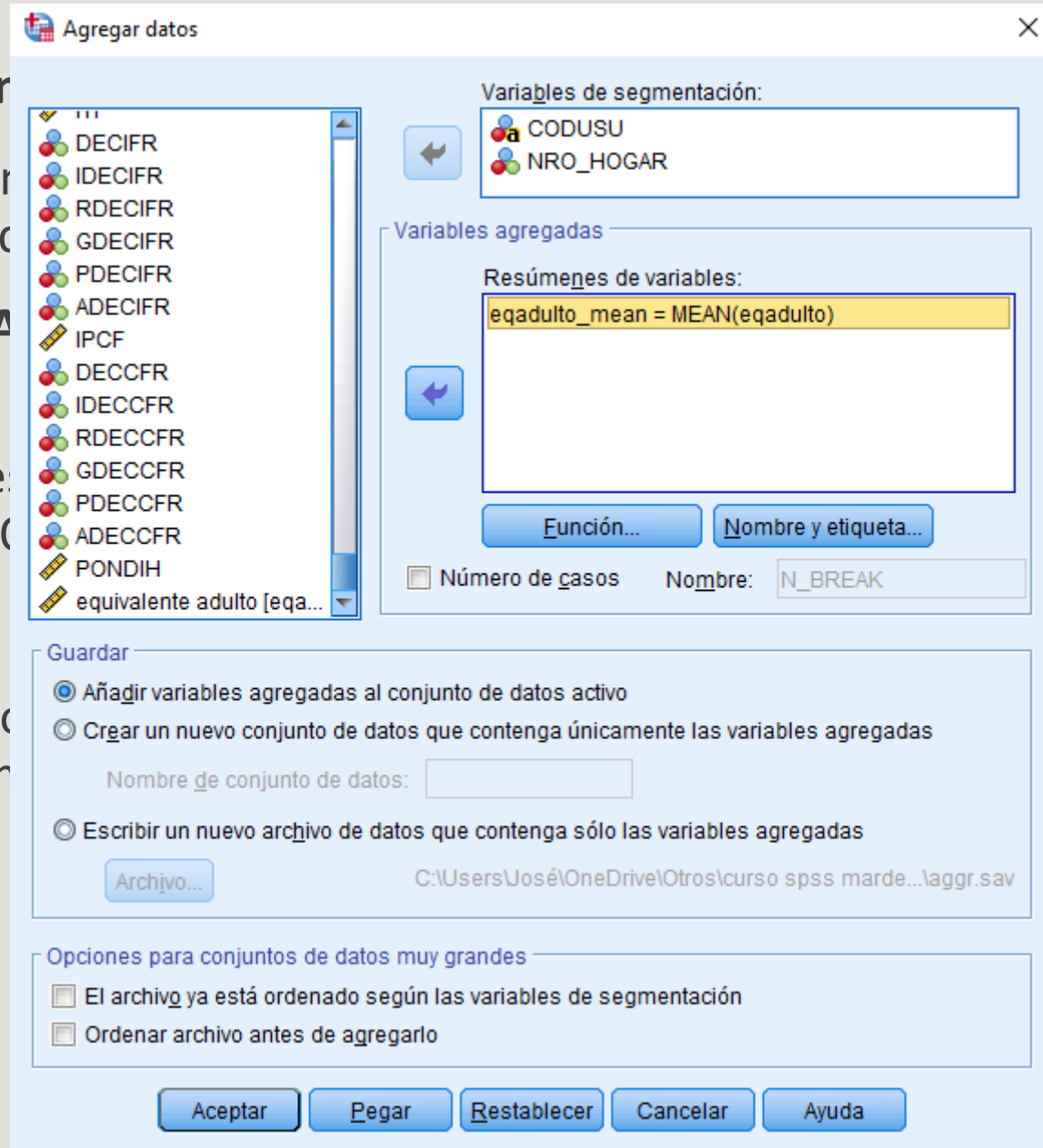
Utilización del comando “Comando de agregación de datos” para crear una variable en función de su edad.

2) Utilización del comando Agregar datos a partir de valores en cada hogar.

Dicho comando sirve para crear una nueva variable de segmentación (CODUSU) a partir de los datos de los hogares.

Datos -> Agregar

Se crea una nueva variable con el nombre N_BREAK que puede ser agregada a la misma base de datos.



3) Abrir la base de hogares y sumar la variable que contiene la sumatoria en la medida de Adulto Equivalente

Dicho proceso se hace con el comando “Fusionar archivos”

Datos -> Fusionar archivos -> Añadir variables

Se toman los datos de la base en la que guardamos la variable de “Eqadulto” por hogar

Lo importante en este caso es identificar las variable que funcionan como claves (CODUSU y Nro_hogar)

El proceso pega la variable en la base de hogares

	ADECCFR	PONDIH	VII1_1	VII1_2	VII2_1	VII2_2	VII2_3	VII2_4	eqadulto_sum
1	9	158	2	0	98	0	0	0	1,50
2	12	0	1	0	98	0	0	0	,67
3	3	504	1	2	3	4	0	0	3,22
4	8	2300	1	2	98	0	0	0	1,77
5	2	331	2	0	98	0	0	0	3,43
6	4	464	2	3	98	0	0	0	5,14
7	8	170	1	2	98	0	0	0	1,76

Fusión base de datos de adulto equivalente a base hogar (1:1) – Versión SPSS 25

Añadir variables de C:\Users\José\OneDrive\Otros\curso spss mardel\... X

Método de fusión Variables

Fusión de uno con uno basada en el orden de archivo

Fusión de uno con uno basada en valores de clave

Fusión de uno con muchos basada en valores de clave

Seleccionar tabla de búsqueda

ConjuntoDatos4*

aggr.sav

*Conjunto de datos activos

i Para una fusión basada en valores de clave, los archivos se deben ordenar en el orden de los valores de clave

Ordenar archivos por valores de clave antes de fusionar

Variables clave:

CODUSU

NRO_HOGAR

i Utilice la pestaña Variables para añadir o eliminar variables de clave

Aceptar Pegar Restablecer Cancelar Ayuda

Añadir variables de C:\Users\José\OneDrive\Otros\curso spss mardel\... X

Método de fusión Variables

Variables excluidas:

Variables incluidas:

- ANO4(*)
- TRIMESTRE(*)
- REALIZADA(*)
- REGION(*)
- MAS_500(*)
- AGLOMERADO(*)
- PONDERA(*)

Variables clave:

- CODUSU
- NRO_HOGAR

Cambiar nombre...

i Los nombres de variable de clave y el tipo de datos deben ser los mismos en ambos archivos. Para variables que no son de clave, los nombres de variable no pueden ser iguales en ambos archivos. Utilice Cambiar de nombre para cambiar nombres de variable.

(*)=ConjuntoDatos4
(+)=aggr.sav

Aceptar Pegar Restablecer Cancelar Ayuda

Fusión base de datos de adulto equivalente a base hogar (1:1) – Versión SPSS 23

Añadir variables de C:\Users\José\OneDrive\Otros\curso spss mardel\bases\EPH_usu_2_Trim_2017_txt\aggr.sav

Variables excluidas:

[Cambiar nombre...](#)

Asignar casos en variables clave

Los casos están ordenados por orden de variables clave en ambos conjuntos de datos

El que no es conjunto de datos activo es una tabla de claves

El conjunto de datos activo es una tabla de claves

Am**o**s archivos proporcionan casos

Indicar origen del caso como variable:

(*)=Conjunto de datos activo
(+)=C:\Users\José\OneDrive\Otros\curso spss mardel\bases\EPH_usu_2_Trim_2017_txt\aggr.sav

[Aceptar](#) [Pegar](#) [Restablecer](#) [Cancelar](#) [Ayuda](#)

Nuevo conjunto de datos activo:

ANO4(*)
TRIMESTRE(*)
REALIZADA(*)
REGION(*)
MAS_500<(*)
AGLOMERADO(*)
PONDERA(*)
IV1(*)
IV1_ESP(*)
NRO_HOGAR(*)

Variables clave:

CODUSU>
NRO_HOGAR

4) En función de los valores regionales de la CBA y la CBT, creo las variables que identifican a quienes están o no bajo la línea de pobreza e indigencia

Utilización de la variable “Ingreso Total Familiar” (ITF) como medida de los ingresos alcanzados por el hogar.

Ejemplo región Gran Buenos Aires:

```
Do if (region=1).
```

```
compute li = 1936.38 * eqadulto_sum.
```

```
compute lp = (1936.38 * 2.45) * eqadulto_sum.
```

```
end if.
```

```
if itf < lp pobreza = 1.
```

```
if itf >= lp pobreza = 2.
```

```
if itf < li indigencia = 1.
```

```
if itf >= li indigencia = 2.
```

Uso del comando “Do if”

5) Ya se tiene clasificado a los hogares en función de su posición respecto a la línea de pobreza e indigencia

Para calcular frecuencias siempre activar el ponderador de ingresos de hogar (PONDIH)

Línea de Pobreza

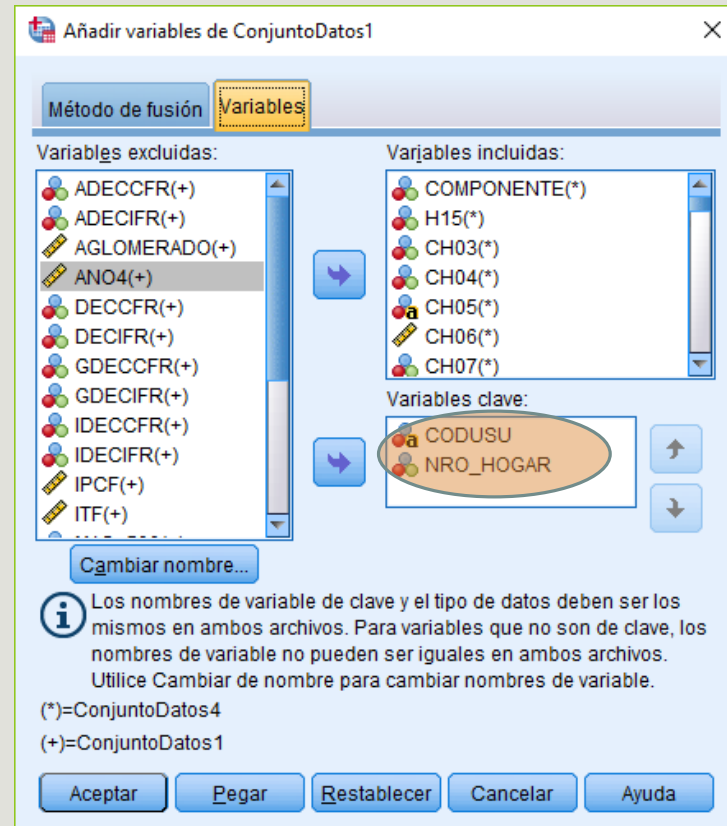
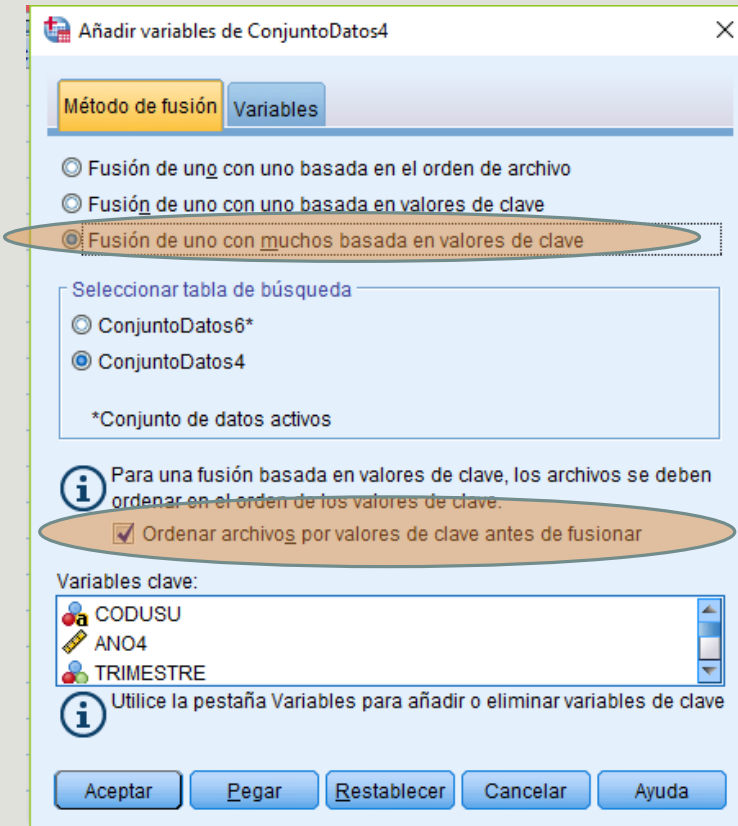
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido Pobre	1861539	21,0	21,0	21,0
No pobre	6996972	79,0	79,0	100,0
Total	8858511	100,0	100,0	

Línea de Indigencia

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido Indigente	389766	4,4	4,4	4,4
No indigente	8468745	95,6	95,6	100,0
Total	8858511	100,0	100,0	

6) Para obtener el porcentaje de individuos bajo la línea de pobreza e indigencia debo pasar las variables “pobreza” e “indigencia” a la base de individuos.

Fusión base de datos de hogar a individuos (muchos:1) – Versión SPSS 25



Fusión base de datos de hogar a individuos (muchos:1) – Versión SPSS 23

Añadir variables de ConjuntoDatos1

Variables excluidas:

- IDECCFR(+)
- IDECIFR(+)
- IPCF(+)
- ITF(+)
- MAS_500<(+)
- PDECCFR(+)
- PDECIFR(+)

[Cambiar nombre...](#)

Asignar casos en variables clave

Los casos están ordenados por orden de variables clave en ambos conjuntos de datos

El que no es conjunto de datos activo es una tabla de claves

El conjunto de datos activo es una tabla de claves

Ambos archivos proporcionan casos

Indicar origen del caso como variable:

(*)=Conjunto de datos activo
(+)=ConjuntoDatos1

[Aceptar](#) [Pegar](#) [Restablecer](#) [Cancelar](#) [Ayuda](#)

Nuevo conjunto de datos activo:

- VII2_1(+)
- VII2_2(+)
- VII2_3(+)
- VII2_4(+)
- eqadulto_sum(+)
- li(+)
- lp(+)
- pobreza(+)
- indigencia(+)

Variables clave:

- CODUSU>
- NRO_HOGAR

6) Para obtener el porcentaje de individuos bajo la línea de pobreza e indigencia debo pasar las variables “pobreza” e “indigencia” a la base de individuos.

Uso el comando “fusionar archivos” nuevamente pero desde la base de individuos.

Debo pegar únicamente las variables de interés (pobreza e indigencia), no así las restantes de la base hogar.

Línea de Pobreza

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido Pobre	8050760	29,3	29,3	29,3
No pobre	19435498	70,7	70,7	100,0
Total	27486258	100,0	100,0	

Línea de Indigencia

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido Indigente	1697651	6,2	6,2	6,2
No indigente	25788607	93,8	93,8	100,0
Total	27486258	100,0	100,0	

Ahora veamos el porcentaje de pobreza por grupos.

- Para esto utilizaremos el comando “tablas personalizadas” que permite crear tablas flexibles con la información que queramos
- Analizar -> Tablas -> Tablas personalizadas

Tablas personalizadas

Tabla Tablas Estadísticos de prueba Opciones

Variables: Normal Compacta Casas

Columnas

		Línea de Pobreza		Línea de Indigencia	
		1,00 Pobre	2,00 No ...	1,00 ...	2,00 No ...
		% del N de ...	% del N de ...	% del N de ...	% del N de ...
Sexo	1 Varón	nnn,n%	nnn,n%	nnn,n%	nnn,n%
	2 Mujer	nnn,n%	nnn,n%	nnn,n%	nnn,n%
Nivel educativo	1 Primaria ...	nnn,n%	nnn,n%	nnn,n%	nnn,n%
	2 Primaria ...	nnn,n%	nnn,n%	nnn,n%	nnn,n%
	3 ...	nnn,n%	nnn,n%	nnn,n%	nnn,n%
	4 ...	nnn,n%	nnn,n%	nnn,n%	nnn,n%
	5 Superior ...	nnn,n%	nnn,n%	nnn,n%	nnn,n%
	6 Superior ...	nnn,n%	nnn,n%	nnn,n%	nnn,n%
	7 Sin ...	nnn,n%	nnn,n%	nnn,n%	nnn,n%
Categoría ocupacional	9 Ns J Nr	nnn,n%	nnn,n%	nnn,n%	nnn,n%
	1 Patrón	nnn,n%	nnn,n%	nnn,n%	nnn,n%
	2 Cuenta ...	nnn,n%	nnn,n%	nnn,n%	nnn,n%
	3 Obrero o ...	nnn,n%	nnn,n%	nnn,n%	nnn,n%
Categorías	4 ...	nnn,n%	nnn,n%	nnn,n%	nnn,n%
	9 Ns J Nr	nnn,n%	nnn,n%	nnn,n%	nnn,n%
	1 GBA	nnn,n%	nnn,n%	nnn,n%	nnn,n%
	40 Noroeste	nnn,n%	nnn,n%	nnn,n%	nnn,n%
	41 Noroeste	nnn,n%	nnn,n%	nnn,n%	nnn,n%

Define Estadísticos de resumen

Estadísticos de resumen: Estadísticos de resumen Categorías y Totales

Posición: Columnas Ocultar Posición de categoría: Valor predeterminado

Origen: Variables de columna

Aceptar Pegar Restablecer Cancelar Ayuda

		Línea de Pobreza		Línea de Indigencia	
		1,00 Pobre % del N de fila	2,00 No pobre % del N de fila	1,00 Indigente % del N de fila	2,00 No indigente % del N de fila
Sexo	1 Varón	30,0%	70,0%	6,4%	93,6%
	2 Mujer	28,7%	71,3%	5,9%	94,1%
Nivel educativo	1 Primaria Incompleta	42,2%	57,8%	10,0%	90,0%
	2 Primaria Completa	31,9%	68,1%	6,7%	93,3%
	3 Secundaria Incompleta	39,3%	60,7%	8,8%	91,2%
	4 Secundaria Completa	22,6%	77,4%	3,6%	96,4%
	5 Superior Universitaria Incompleta	16,1%	83,9%	1,6%	98,4%
	6 Superior Universitaria Completa	4,9%	95,1%	0,9%	99,1%
	7 Sin instrucción	41,7%	58,3%	10,5%	89,5%
Categoría ocupacional	0	34,9%	65,1%	8,1%	91,9%
	1 Patrón	11,4%	88,6%	1,5%	98,5%
	2 Cuenta propia	31,6%	68,4%	7,3%	92,7%
	3 Obrero o empleado	20,4%	79,6%	3,0%	97,0%
	4 Trabajador familiar sin remuneración	30,4%	69,6%	8,2%	91,8%
	9 Ns./Nr	55,6%	44,4%	55,6%	44,4%
Region	1 GBA	29,6%	70,4%	7,1%	92,9%
	40 Noroeste	29,6%	70,4%	5,0%	95,0%
	41 Noreste	33,1%	66,9%	4,9%	95,1%
	42 Cuyo	27,1%	72,9%	4,0%	96,0%
	43 Pampeana	29,1%	70,9%	5,7%	94,3%
	44 Patagónica	23,1%	76,9%	3,4%	96,6%
	Total	29,3%	70,7%	6,2%	93,8%

		Línea de Pobreza		Línea de Indigencia	
		1,00 Pobre	2,00 No pobre	1,00 Indigente	2,00 No indigente
		% del N de fila	% del N de fila	% del N de fila	% del N de fila
AGLOMERADO	2 Gran La Plata	32,5%	67,5%	5,4%	94,6%
	3 Bahía Blanca - Cerri	20,3%	79,7%	3,3%	96,7%
	4 Gran Rosario	27,0%	73,0%	6,1%	93,9%
	5 Gran Santa Fé	29,8%	70,2%	3,3%	96,7%
	6 Gran Paraná	24,1%	75,9%	6,3%	93,7%
	7 Posadas	30,4%	69,6%	3,2%	96,8%
	8 Gran Resistencia	34,3%	65,7%	3,6%	96,4%
	9 Cdro. Rivadavia - R.Tilly	20,4%	79,6%	3,5%	96,5%
	10 Gran Mendoza	32,1%	67,9%	4,9%	95,1%
	12 Corrientes	38,8%	61,2%	8,1%	91,9%
	13 Gran Córdoba	31,9%	68,1%	6,6%	93,4%
	14 Concordia	45,0%	55,0%	6,7%	93,3%
	15 Formosa	26,3%	73,7%	4,4%	95,6%
	17 Neuquén - Plottier	29,5%	70,5%	3,9%	96,1%
	18 S.del Estero - La Banda	45,9%	54,1%	14,0%	86,0%
	19 Jujuy - Palpalá	24,1%	75,9%	1,0%	99,0%
	20 Río Gallegos	10,8%	89,2%	0,8%	99,2%
	22 Gran Catamarca	30,3%	69,7%	4,5%	95,5%
	23 Salta	30,6%	69,4%	5,5%	94,5%
	25 La Rioja	21,4%	78,6%	0,3%	99,7%
	26 San Luis - El Chorrillo	25,0%	75,0%	1,4%	98,6%
	27 Gran San Juan	18,6%	81,4%	3,5%	96,5%
	29 Gran Tucumán - T. Viejo	25,7%	74,3%	3,6%	96,4%
	30 Santa Rosa - Toay	29,8%	70,2%	2,1%	97,9%
	31 Ushuaia - Río Grande	12,6%	87,4%	2,4%	97,6%
	32 Ciudad de Bs As	14,2%	85,8%	2,7%	97,3%
	33 Partidos del GBA	33,4%	66,6%	8,2%	91,8%
	34 Mar del Plata - Batán	25,1%	74,9%	6,8%	93,2%
	36 Río Cuarto	25,9%	74,1%	4,5%	95,5%
	38 San Nicolás - Villa Constitución	30,8%	69,2%	6,9%	93,1%
	91 Rawson - Trelew	29,1%	70,9%	3,8%	96,2%
	93 Viedma - Carmen de Patagones	32,8%	67,2%	6,0%	94,0%

Datos para el aglomerado Mar del Plata - Batán

		Línea de Pobreza		Línea de Indigencia	
		1,00 Pobre % del N de fila	2,00 No pobre % del N de fila	1,00 Indigente % del N de fila	2,00 No indigente % del N de fila
Sexo	1 Varón	25,7%	74,3%	5,5%	94,5%
	2 Mujer	24,6%	75,4%	7,9%	92,1%
Nivel educativo	1 Primaria Incompleta	35,6%	64,4%	15,2%	84,8%
	2 Primaria Completa	28,7%	71,3%	7,5%	92,5%
	3 Secundaria Incompleta	35,0%	65,0%	6,2%	93,8%
	4 Secundaria Completa	14,0%	86,0%	3,4%	96,6%
	5 Superior Universitaria Incompleta	15,5%	84,5%	0,0%	100,0%
	6 Superior Universitaria Completa	3,2%	96,8%	0,8%	99,2%
	7 Sin instrucción	43,6%	56,4%	17,7%	82,3%
Categoría ocupacional	0	29,4%	70,6%	9,0%	91,0%
	1 Patrón	16,5%	83,5%	6,9%	93,1%
	2 Cuenta propia	23,1%	76,9%	4,0%	96,0%
	3 Obrero o empleado	17,9%	82,1%	3,9%	96,1%
	4 Trabajador familiar sin remuneración	100,0%	0,0%	0,0%	100,0%
	Total	25,1%	74,9%	6,8%	93,2%

Brechas de pobreza e indigencia

- Miden la distancia entre el ingreso total familiar promedio de los pobres e indigentes y el promedio de las líneas de pobreza e indigencia.

- Ejemplo brecha de pobreza:

ITF prom. hogares pobres – Línea de pobreza prom. hogares pobres

- Para calcular cada brecha, debemos seleccionar a los casos que están bajo ambas líneas.
- En SPSS para calcular las medias, podemos hacerlo con el comando “descriptivos”
- Analizar -> estadísticos descriptivos -> descriptivos

Estadísticos descriptivos

	N	Media
Línea de Pobreza por hogar	1861539	15861,1894
ITF	1861539	9954,92
N válido (por lista)	1861539	

-5906,27

Brecha monetaria

37,2%

Brecha porcentual

Estadísticos descriptivos

	N	Media
Línea de Indigencia por hogar	389766	6523,9009
ITF	389766	3936,89
N válido (por lista)	389766	

-2587,01

39,6%

Introducción al estudio de la desigualdad social a partir del programa STATA

JOSE RODRÍGUEZ DE LA FUENTE (IIGG – UBA –
CONICET – INCASI)

Propuesta de la sesión

- Introducir características básicas y centrales del programa STATA
- Elaboración de brechas de ingresos
- Cálculo del coeficiente de GINI y THEIL
- Descomposición de los coeficientes

¿Por qué STATA? ¿Por qué estudiar la desigualdad social con STATA?

	Potencia	Dificultad de uso	Precio	Uso prioritario
SPSS	Alta	Baja*	Muy alto	Docencia / profesional
Stata	Muy alta	Media	Medio/Alto	Investigación aplicada
R	Máxima	Alta	Gratuito	Investigación básica y aplicada

Fuente: Curso de introducción a Stata – Jordi Muñoz (UAB)

¿Cómo usar STATA?

- Interfaz
- Línea de comando
- Sintaxis (do-files)

- Automatización de tareas complejas
- Generar comandos propios (ej: gini)
- Capacidad de reproducir manipulaciones de datos y resultados (¡lo más importante!)



#	Command	_rc
1	findit fastgini	
2	findit fastgini	
3	findit ineqerr	
4	use "C:\Users\José\...	
5	describe region	
6	codebook region	
7	sum region	

```
Statistics/Data Analysis
MP - Parallel Edition
13.0 Copyright 1985-2013 StataCorp LP
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC http://www.stata.com
979-696-4600 stata@stata.com
979-696-4601 (fax)

3-user 8-core Stata network perpetual license:
Serial number: 501306208483
Licensed to: IDRE-UCLA
IDRE-UCLA

Notes:
1. (/v# option or -set maxvar-) 500...

. findit fastgini
. findit fastgini
. findit ineqerr
. use "C:\Users\José\OneDrive\Bases\EPH\STATA\Individual_t207.dta", clear
. describe region

      storage   display   value
variable name  type   format   label   variable label
-----
region         double %10.0g  region  Código de Región
```

Variable	Label
CODUSU	Código para distin...
nro_hogar	Código para distin...
componente	Número de comp...
h15	Entrevista individu...
ano4	Año de relevamie...
	mes...
	gión
	seg...
aglomerado	Agglomerado
pondera	Ponderación
ch03	Relación de paren...
ch04	Sexo
ch06	Edad en años cum...

Name	region
Label	Código de Región
Type	double
Form	
Valu	
Not	

Property	Value
Variables	176
Observations	63.235
Size	83,70M

Data editor (base)

Do-file

Ventana de revisión

Ventana de resultados

Ventana de comando

Ventana de variables

Ventana de propiedades de variables

Data editor (base)

	CODUSU	nro_hogar	componente	h15	ano4	trimestre	region	mas_500	aglomerado	pondera	ch03	ch04	ch04
1	125009	1	2	Sí	2007	2do. Trime	Gran Bueno	S	Ciudad de	931	Jefe	Varón	
2	125009	1	3	Sí	2007	2do. Trime	Gran Bueno	S	Ciudad de	931	Hermano	Mujer	
3	125009	1	4	Sí	2007	2do. Trime	Gran Bueno	S	Ciudad de	931	Hermano	Varón	
4	125009	1	5	Sí	2007	2do. Trime	Gran Bueno	S	Ciudad de	931	Hermano	Mujer	
5	125250	1	1	Sí	2007	2do. Trime	Gran Bueno	S	Ciudad de	1408	Jefe	Varón	
6	125250	1	2	Sí	2007	2do. Trime	Gran Bueno	S	Ciudad de	1408	Cónyuge/Pa	Mujer	
7	125250	1	3	0	2007	2do. Trime	Gran Bueno	S	Ciudad de	1408	Hijo/Hijas	Mujer	
8	125250	1	4	Sí	2007	2do. Trime	Gran Bueno	S	Ciudad de	1408	Suegro	Mujer	
9	125963	1	1	Sí	2007	2do. Trime	Gran Bueno	S	Ciudad de	1154	Jefe	Mujer	
10	125963	1	2	Sí	2007	2do. Trime	Gran Bueno	S	Ciudad de	1154	Hijo/Hijas	Mujer	
11	125963	1	3	Sí	2007	2do. Trime	Gran Bueno	S	Ciudad de	1154	Madre/Padr	Mujer	
12	126002	1	1	Sí	2007	2do. Trime	Gran Bueno	S	Ciudad de	457	Jefe	Varón	
13	126002	1	2	Sí	2007	2do. Trime	Gran Bueno	S	Ciudad de	457	Cónyuge/Pa	Mujer	
14	126002	1	3	Sí	2007	2do. Trime	Gran Bueno	S	Ciudad de	457	Hijo/Hijas	Varón	
15	126002	1	4	Sí	2007	2do. Trime	Gran Bueno	S	Ciudad de	457	Hijo/Hijas	Varón	
16	126002	2	servicio d	Sí	2007	2do. Trime	Gran Bueno	S	Ciudad de	457	Jefe	Mujer	
17	126289	1	1	Sí	2007	2do. Trime	Gran Bueno	S	Ciudad de	1393	Jefe	Mujer	
18	126289	2	servicio d	Sí	2007	2do. Trime	Gran Bueno	S	Ciudad de	1393	Jefe	Mujer	
19	126442	1	1	Sí	2007	2do. Trime	Gran Bueno	S	Ciudad de	1834	Jefe	Varón	
20	126442	1	2	Sí	2007	2do. Trime	Gran Bueno	S	Ciudad de	1834	Cónyuge/Pa	Mujer	
21	126467	1	1	Sí	2007	2do. Trime	Gran Bueno	S	Ciudad de	1767	Jefe	Mujer	
22	126467	1	2	Sí	2007	2do. Trime	Gran Bueno	S	Ciudad de	1767	Cónyuge/Pa	Varón	
23	126467	1	3	Sí	2007	2do. Trime	Gran Bueno	S	Ciudad de	1767	Hijo/Hijas	Mujer	
24	126467	1	4	Sí	2007	2do. Trime	Gran Bueno	S	Ciudad de	1767	Hijo/Hijas	Mujer	
25	126683	1	1	Sí	2007	2do. Trime	Gran Bueno	S	Ciudad de	1569	Jefe	Varón	
26	126683	1	2	Sí	2007	2do. Trime	Gran Bueno	S	Ciudad de	1569	Cónyuge/Pa	Mujer	
27	126683	1	3	0	2007	2do. Trime	Gran Bueno	S	Ciudad de	1569	Hijo/Hijas	Mujer	
28	126716	1	1	Sí	2007	2do. Trime	Gran Bueno	S	Ciudad de	1568	Jefe	Varón	

Variables

Filter variables here

Variable	Label	
<input checked="" type="checkbox"/>	CODUSU	Código para distin...
<input checked="" type="checkbox"/>	nro_hogar	Código para distin...
<input checked="" type="checkbox"/>	componente	Número de comp...
<input checked="" type="checkbox"/>	h15	Entrevista individu...
<input checked="" type="checkbox"/>	ano4	Año de relevamie...
<input checked="" type="checkbox"/>	trimestre	Número de trimes...
<input checked="" type="checkbox"/>	region	Código de Región
<input checked="" type="checkbox"/>	mas_500	Aglomerados seg...
<input checked="" type="checkbox"/>	aglomerado	Aglomerado
<input checked="" type="checkbox"/>	pondera	Ponderación
<input checked="" type="checkbox"/>	ch03	Relación de paren...
<input checked="" type="checkbox"/>	ch04	Sexo

Properties

Variables

Name	CODUSU
Label	Código para disti...
Type	str8
Format	%8s
Value Label	
Notes	

Data

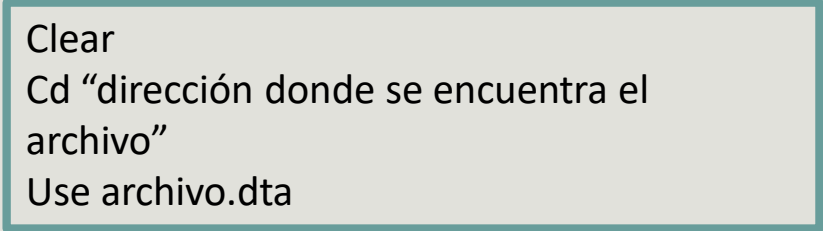
Filename	Individual_t207.dta
Label	
Notes	
Variables	176
Observations	63,235
Size	83,70M
Memory	128M

Funciones de ayuda de STATA

- Elemento central que provee STATA
- Comando “**help**”: brinda información completa acerca del comando sobre el que tenemos dudas (descripción, opciones, ejemplos, videos, etc.)
 - Sintaxis: help [nombre del comand] (ej: help tab)
- Comando “**Findit**”: buscar programas (rutinas) creadas por otros usuarios para agregar al STATA (ej: gini, multiple correspondence)
- Importante **comunidad de usuarios** que consultan y responden a interrogantes en tono al programa (foros, blogs, videos, STATALIST, página UCLA, etc.)

Abriendo la base de datos

- Formatos que soporta STATA: dta, xls, csv, txt, etc.
- Si queremos importar archivos de .sav, debemos primero guardarlos como .dta en el SPSS, para luego abrirlo con el STATA.
- Opciones para abrir una base:
 - A través de la interfaz grafica
 - Con sintaxis



```
Clear  
Cd "dirección donde se encuentra el  
archivo"  
Use archivo.dta
```


Lenguaje STATA

- comando [variable/s], [opciones]
- se puede abreviar (ej: summarize/sum; tabulate/tab)
- agregar comentarios en sintaxis (*), agregar comentarios en una línea donde hay órdenes (//)
- ¡Ojo! Hay confusiones con respecto al (=). El (=) funciona como operador aritmético y el (==) como operador lógico (prueba de equidad).

Inspeccionando las variables (comandos)

- DESCRIBE: brinda información sobre el tipo de variable que seleccionemos
- SUMMARIZE: información básica sobre como se distribuyen los casos según la variable (obs, media, desvío estándar, min, max)
- CODEBOOK: es como el summarize pero brinda mayor información

Distribución de frecuencias

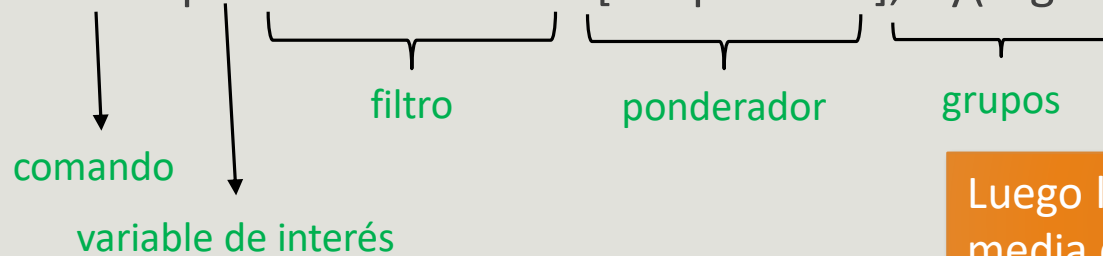
- **Comando TAB:** nos permite ver como se distribuyen las frecuencias según variable.
- También permite realizar tablas de contingencias de dos o más variables.
- Permite el uso de “condicionantes” (IF) o “segmentaciones” (BYSORT).

Brechas de ingresos

- El concepto de brecha remite a la idea de distancia.
- Generalmente, las brechas se calculan en referencia al valor medio de la variable de interés.
- Ej: ingresos, horas de trabajo, etc.
- Para realizar esta operación en STATA, debemos utilizar el comando TABSTAT.
- El mismo permite crear tablas que resumen determinados estadísticos (media, desvío, mediana, varianza, etc.)

Sintaxis:

```
tabstat p47t if estado == 1 [fw=pondera], by(region)
```



Luego lo pego en Excel y lo calculo:
media del grupo / media total

```
Summary for variables: p47t  
by categories of: region (Código de Región)
```

region	mean
Gran Buenos Aire	8116,561
NOA	6374,475
NEA	6128,378
Cuyo	7432,952
Pampeana	7948,285
Patagónica	12076,58
Total	7930,47

region	mean	brecha
Buenos Aire	8116,56	1,02
NOA	6374,48	0,80
NEA	6128,38	0,77
Cuyo	7432,95	0,94
Pampeana	7948,29	1,00
Patagónica	12076,58	1,52
Total	7930,47	1,00

El análisis de la desigualdad a partir del coeficiente de GINI

- ¿Qué necesito conocer?
 - Frecuencia acumulada de ingresos
 - Frecuencia acumulada de personas
- Es una medida de dispersión
 - 1 → Máxima desigualdad
 - 0 → Máxima igualdad
- La definición mas corriente se formula en términos de la curva de Lorenz
- Puede ser calculado a partir de información agregada o individual.
- Puede ser descompuesto algebraicamente.



Comandos de STATA (GINI)

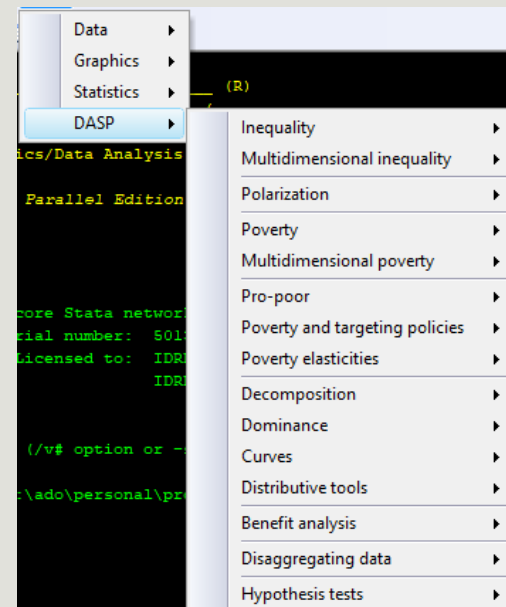
- Buscaremos la palabra GINI en el repositorio de STATA (findit o search)
- Buscaremos los siguientes programas específicos e instalarlos:
 - Fastgini
 - Ineqerr
 - Inequal
 - Rspread
- Calcularemos el índice para el IPCF. Revisemos la particularidad de cada uno de los programas y si presentan diferencias en el cálculo.

Comandos de STATA (GINI)

Programa	Características
fastgini	Es rápido Por default, elimina los casos \neq a 0
rspread	Calcula varias medidas de dispersión Permite graficar la curva de Lorenz
ineqerr	Es más lento Calcula el GINI, THEIL y Var Logs Calcula los errores standard
Inequal	Brinda diversas medidas de dispersión

Descomposición GINI y THEIL

- DASP (**D**istributive **A**nalysis **S**tata **P**ackage) es un paquete que permite analizar la distribución de las condiciones de vida de la población.
- Tiene los principales índices de desigualdad utilizados internacionalmente.



Descomposición GINI y THEIL

- Se instala a partir de los archivos descargados desde la página <http://dasp.ecn.ulaval.ca/index.html>
- Procedimiento:
 - Descomprimir DASP_V2.3.rar y pegar la carpeta DASP en el C:
 - Correr la sintaxis de “install_dasp2.3” en Stata
 - Copiar archivo profile.do y pegar en carpeta /ado/personal

Descomposición GINI y THEIL

- Esta técnica nos permite descomponer al coeficiente de Gini en tres elementos:
 - 1) la parte atribuible a la desigualdad de ingreso dentro de cada estrato (INTRA –WITHIN);
 - 2) la parte atribuible a la desigualdad de ingreso entre cada estrato (INTER – BETWEEN) y
 - 3) un componente residual llamado “término de interacción”, que representa el grado de superposición o traslape de la distribución del ingreso (OVERLAP).

Descomposición GINI y THEIL

- Comando DIGINIG: descomposición del índice de GINI por grupos
- Primero es necesario declarar el tipo de encuesta
 - svyset [weight=pondera], vce(linearized) singleunit(missing)
- User -> DASP -> Decomposition -> Gini: decomposition by groups

The screenshot shows a dialog box titled "DASP|Decomposition of the Gini index of inequality by groups --> diginig command". The dialog has two tabs: "Main" and "Results". In the "Main" tab, there are three dropdown menus for "Variable of interest:" (set to "ipcf"), "Size variable:" (empty), and "Group variable:" (set to "region"). To the right, there is a section for "Index:" with a dropdown menu set to "Gini Index". At the bottom right of the dialog is a "Survey settings..." button. At the bottom of the window are three buttons: "OK", "Cancel", and "Submit".

Decomposition of the Gini Index by Groups

Sampling weight : pondera

Group variable : region

Group	Gini index	Population Share	Income Share	Absolute Contrib.	Relative Contrib.
1: Gran Buenos Aires	0,419	0,513	0,525	0,113	0,273
	0,006	0,005	0,007	0,003	0,007
40: NOA	0,398	0,084	0,066	0,002	0,005
	0,008	0,002	0,002	0,000	0,000
41: NEA	0,389	0,048	0,033	0,001	0,002
	0,007	0,001	0,001	0,000	0,000
42: Cuyo	0,402	0,063	0,059	0,001	0,004
	0,012	0,002	0,002	0,000	0,000
43: Pampeana	0,390	0,252	0,259	0,025	0,062
	0,005	0,004	0,005	0,001	0,002
44: Patagónica	0,411	0,039	0,057	0,001	0,002
	0,007	0,001	0,002	0,000	0,000
Within	---	---	---	0,144	0,347
	---	---	---	---	---
Beetwin	---	---	---	0,054	0,129
	---	---	---	---	---
Overlap	---	---	---	0,217	0,523
	---	---	---	---	---
Population	Gini total	1,000	1,000	0,414	1,000
		0,000	0,000	0,004	0,000

Participación de la desigualdad entre grupos, intra grupos y superposición

Gini de cada región

Distribución de la población

Distribución del ingreso por región

- Comando “dentropyg”: permite descomponer el coeficiente de Theil.
- El mismo presenta mejores propiedades para su descomposición
- No tiene el componente de superposición
- Los pasos son los mismos que con el “diginig”

Group	Entropy index	Population share	$(\mu_k/\mu)^\theta$	Absolute contribution	Relative contribution
1: Gran Buenos Aires	0,296	0,513	1,023	0,155	0,532
	0,011	0,005	0,009	0,006	0,014
40: NOA	0,283	0,085	0,787	0,019	0,065
	0,019	0,002	0,016	0,002	0,005
41: NEA	0,253	0,048	0,683	0,008	0,029
	0,011	0,001	0,014	0,000	0,002
42: Cuyo	0,298	0,063	0,935	0,018	0,061
	0,028	0,002	0,026	0,002	0,007
43: Pampeana	0,256	0,252	1,030	0,066	0,227
	0,010	0,004	0,014	0,003	0,011
44: Patagónica	0,284	0,039	1,466	0,016	0,056
	0,012	0,001	0,030	0,001	0,003
Within	---	---	---	0,283	0,969
	---	---	---	---	---
Between	---	---	---	0,009	0,030
	---	---	---	0,000	---
Population	0,292	1,000	---	0,292	1,000
	0,007	0,000	---	0,007	0,000