

SEMINARIO DE DOCTORADO

**METODOLOGÍA DE
INVESTIGACIÓN SOCIAL**

Agustín Salvia

Santiago Poy

APLICACIÓN

**ANÁLISIS DE TABLAS DE CONTINGENCIA Y
COEFICIENTES DE ASOCIACIÓN**

PARTE 1

Ejercicios – Clase n° 2

1) Construcción de una variable

En función de la base recortada de personas ocupadas de 4to trimestre 2022 para el Gran Buenos Aires, se propone la construcción de un **índice simple de calidad del empleo para asalariados**.

Específicamente, se solicita que construya una variable sumatoria que tome en cuenta los siguientes componentes:

- Si el empleo incluye derecho a vacaciones pagas
- Si el empleo incluye aguinaldo
- Si el trabajador tiene derecho a días pagos por enfermedad
- Si el empleo incluye el pago de obra social
- Si el puesto tiene descuento jubilatorio

Ejercicios – Clase nº 2

1) Construcción de una variable (cont).

```
frequencies PP07G1 PP07G2 PP07G3 PP07G4 PP07H.
```

```
recode PP07G1 PP07G2 PP07G3 PP07G4 PP07H (2=0) (9=0) (1=1)  
(else=sysmis).
```

```
execute .
```

```
value labels PP07G1 PP07G2 PP07G3 PP07G4 PP07H 1'SÍ' 0'No'.
```

```
compute indice_cal=PP07G1+PP07G2+ PP07G3 + PP07G4 + PP07H.
```

```
execute .
```

```
variable labels indice_cal "Indice de calidad del empleo".
```

```
WEIGHT BY PONDERA.
```

```
frequencies indice_cal.
```

Ejercicios – Clase n° 2

2) Caracterice la distribución del ingreso salarial de mujeres y varones en términos de su tendencia central, posición, dispersión absoluta y dispersión relativa.

```
# Varones
USE ALL.
COMPUTE filter_$=(cat_ocup=3 & CH04=1).
FILTER BY filter_$.
EXECUTE.

WEIGHT BY PONDERA.

FREQUENCIES VARIABLES=p21_m
  /NTILES=4
  /STATISTICS=STDDEV VARIANCE MEAN MEDIAN MODE
  /ORDER=ANALYSIS
  /FORMAT=NOTABLE.
```

Ejercicios – Clase n° 2

2) *Caracterice la distribución del ingreso salarial de mujeres y varones en términos de su tendencia central, posición, dispersión absoluta y dispersión relativa.*

p21_m MONTO DE INGRESO DE LA OC

N	Válido	2801822
	Perdidos	0
Media		119175,6778
Mediana		101923,0945
Moda		80000,00
Desv. Desviación		86299,40473
Mínimo		,00
Máximo		1500000,00

N	Válido	2405483
	Perdidos	0
Media		94321,8131
Mediana		84164,2935
Moda		100000,00
Desv. Desviación		79009,53940
Mínimo		,00
Máximo		1080000,00

$$cv = \frac{S}{\bar{x}} = ? \text{ cuál es la distribución más heterogénea?}$$

PARTE 2

Tablas de contingencia en SPSS

SPSS Menú: **Análisis** > **Tablas cruzadas...**

Diálogo: **Tablas cruzadas**

- Filas: info_diario
- Columnas: edad_grupo2

Mostrar en las casillas:

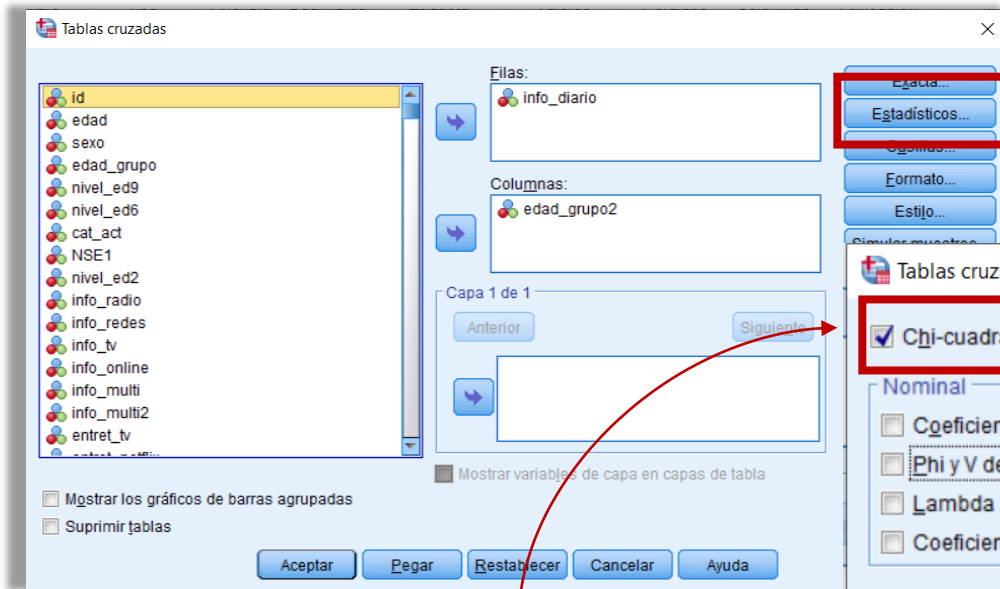
- Recuentos: Observado, Esperado, Ocultar recuentos pequeños (Menor que: 5)
- Prueba z: Comparar las proporciones de columna (Corregir valores p (método de Bonferroni))
- Residuos: No estandarizados, Estandarizados, Estandarizados corregidos
- Porcentajes: Ella, Columna, Total
- Ponderaciones no enteras: Redondear recuentos de casillas, Redondear ponderaciones de casos, Truncar recuentos de casillas, Truncar ponderaciones de casos, No efectuar correcciones

Elegir las variables de filas y columnas

Botón > Casillas

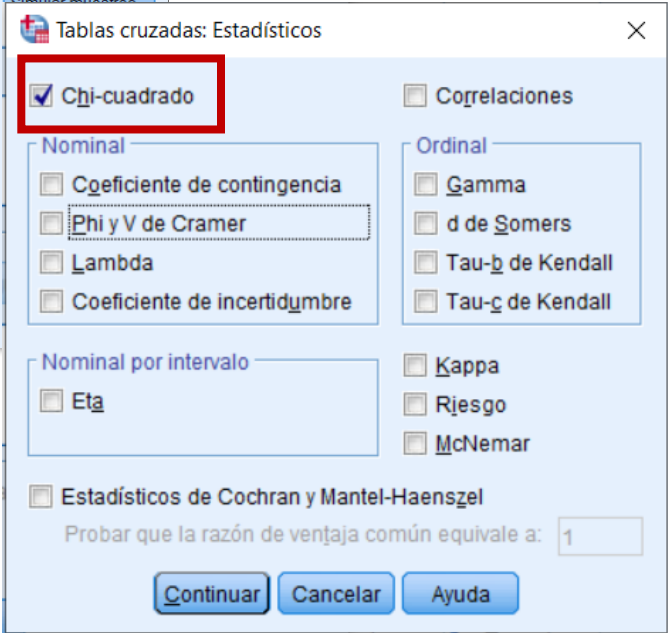
Seleccionamos los porcentajes que queremos

Prueba de independencia estadística χ^2

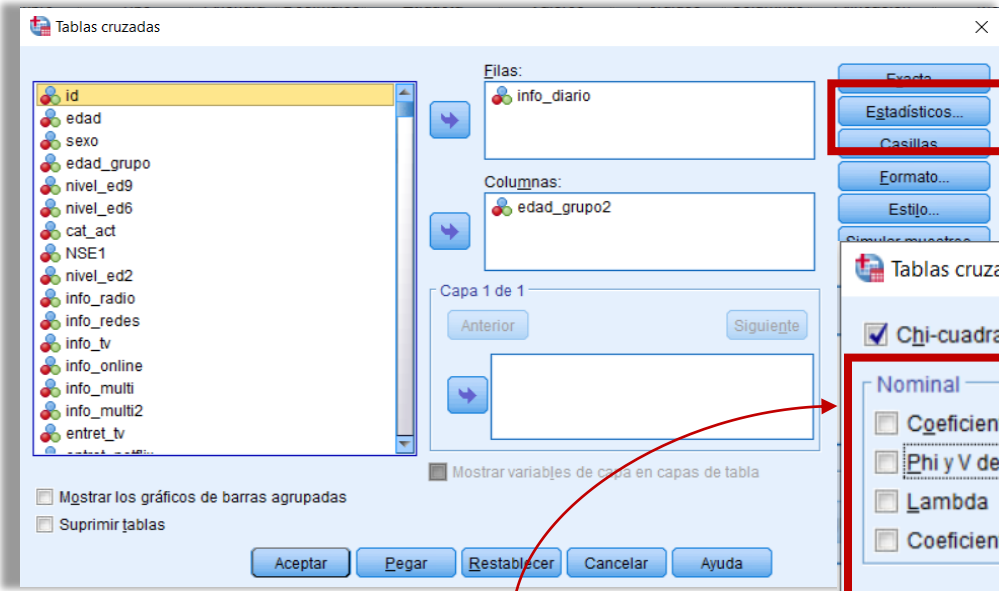


Botón > Estadísticos

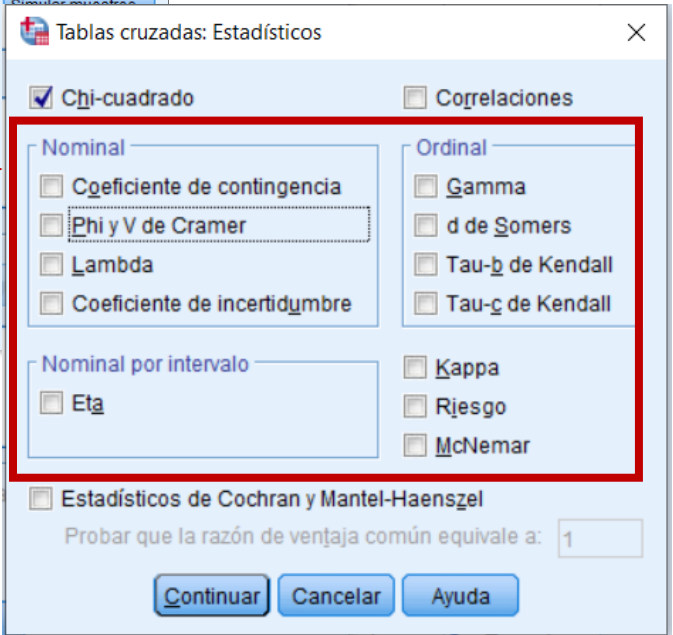
Seleccionamos la prueba



Coefficientes de asociación

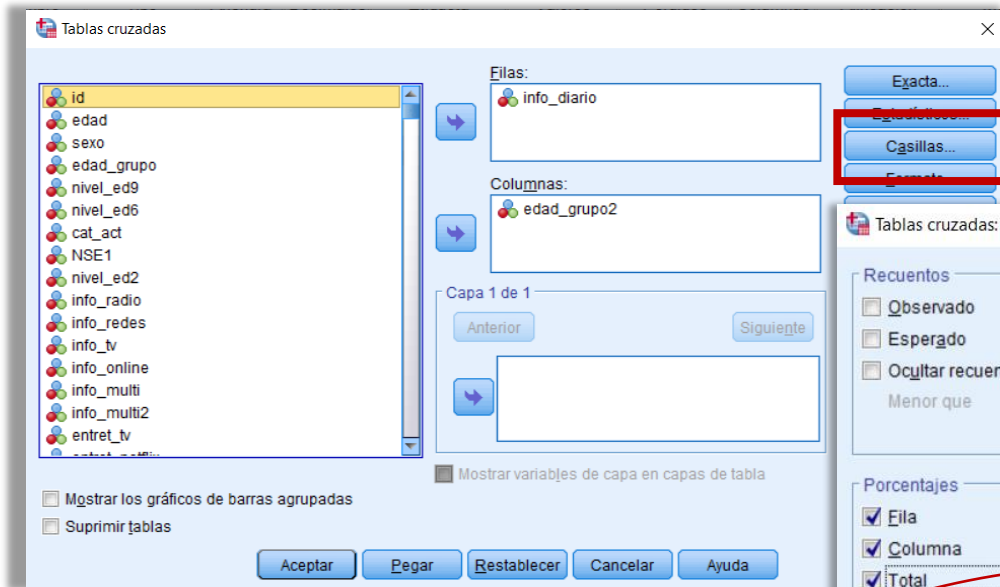


Botón > Estadísticos

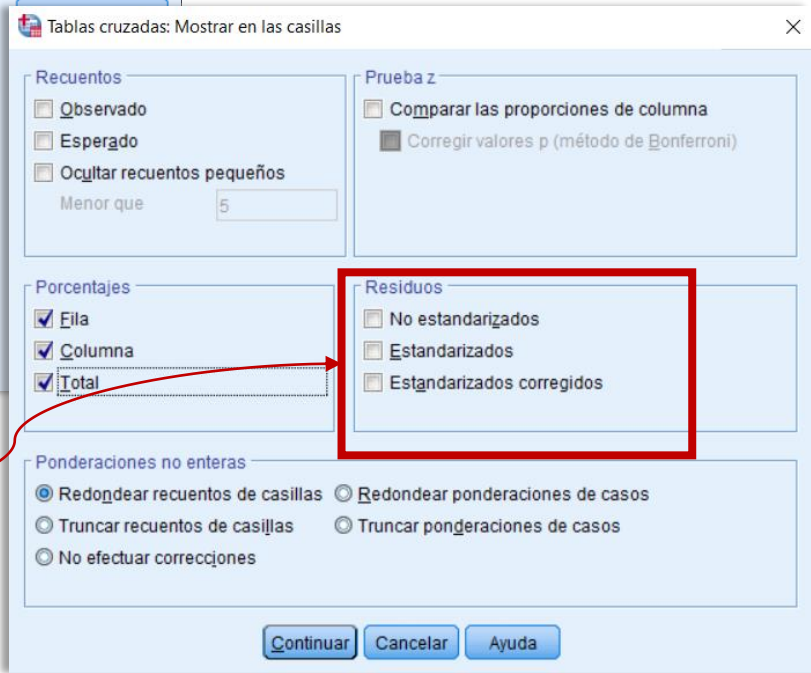


Seleccionamos el coeficiente

Análisis de residuos tipificados

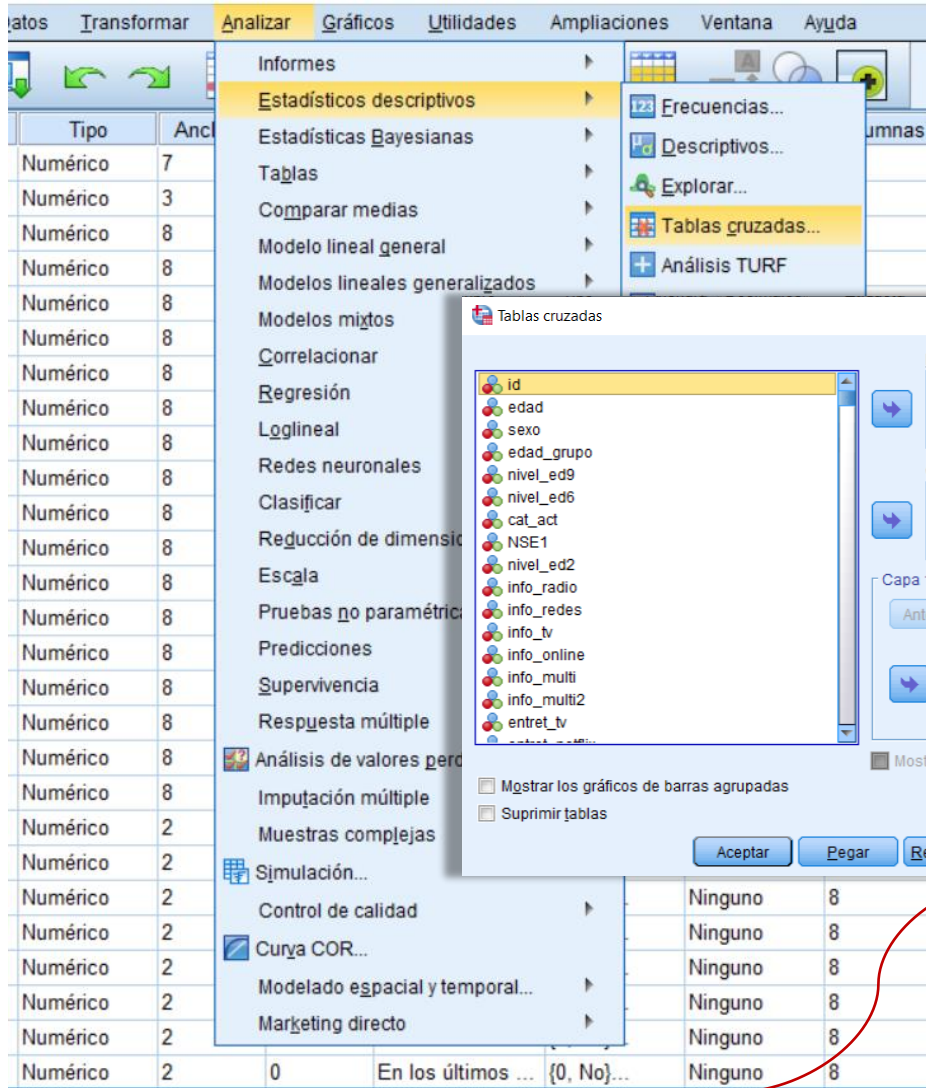


Botón > Casillas



Seleccionamos los residuos

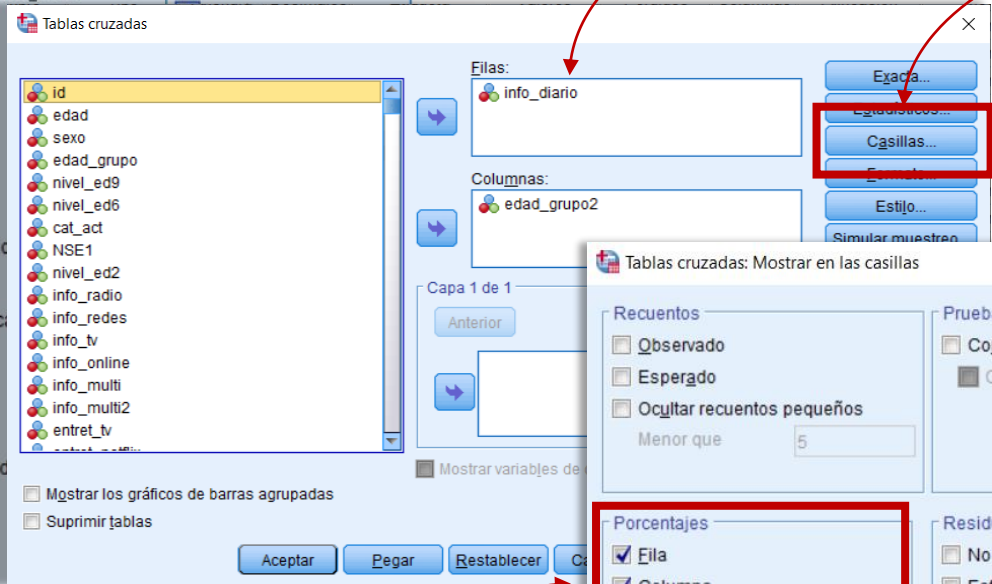
Tablas de contingencia en SPSS



The screenshot shows the SPSS interface with the 'Análisis' menu open. The 'Tablas cruzadas...' option is highlighted. Below the menu, a list of variables is visible, including 'id', 'edad', 'sexo', 'edad_grupo', 'nivel_ed9', 'nivel_ed6', 'cat_act', 'NSE1', 'nivel_ed2', 'info_radio', 'info_redes', 'info_tv', 'info_online', 'info_multi', 'info_multi2', and 'entret_tv'. A small table at the bottom right shows a contingency table with 'Ninguno' as a category and a count of 8.

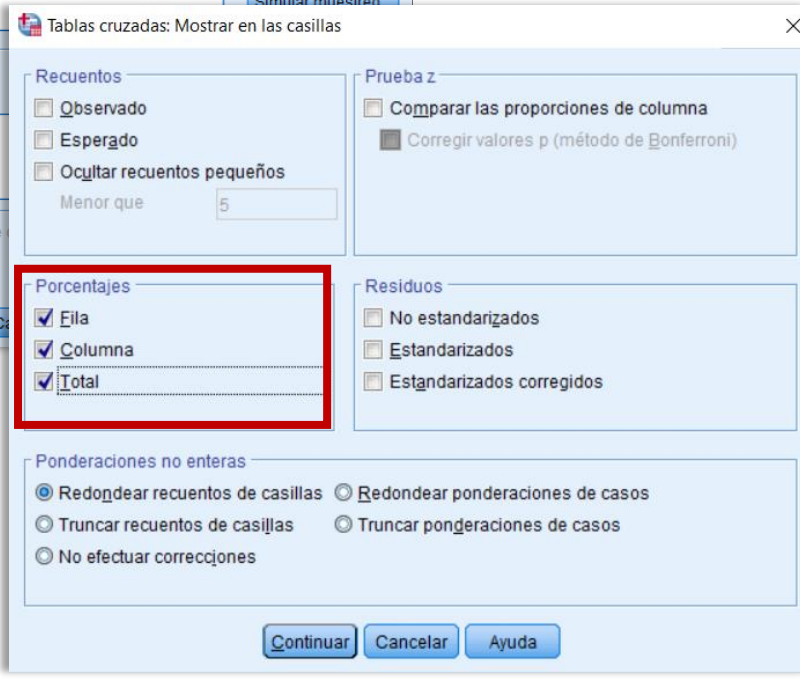
Elegir las variables de filas y columnas

Botón > Casillas



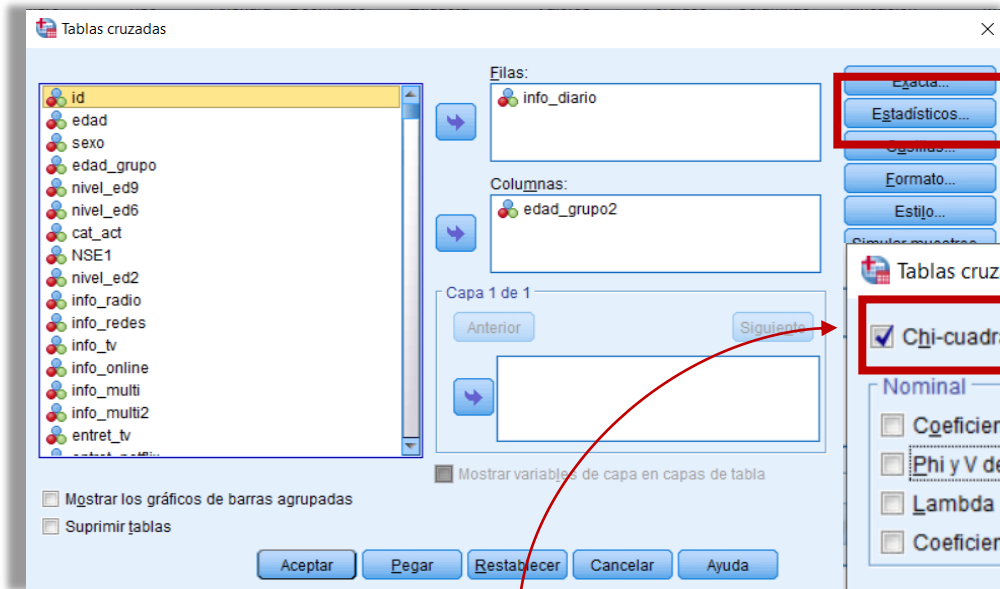
The 'Tablas cruzadas' dialog box is shown. The 'Filas' field contains 'info_diario' and the 'Columnas' field contains 'edad_grupo2'. The 'Casillas...' button is highlighted with a red box. Below the dialog, the 'Mostrar en las casillas' sub-dialog is visible, with the 'Porcentajes' section highlighted by a red box. The 'Porcentajes' section has three checked options: 'Ella', 'Columna', and 'Total'.

Seleccionamos los porcentajes que queremos



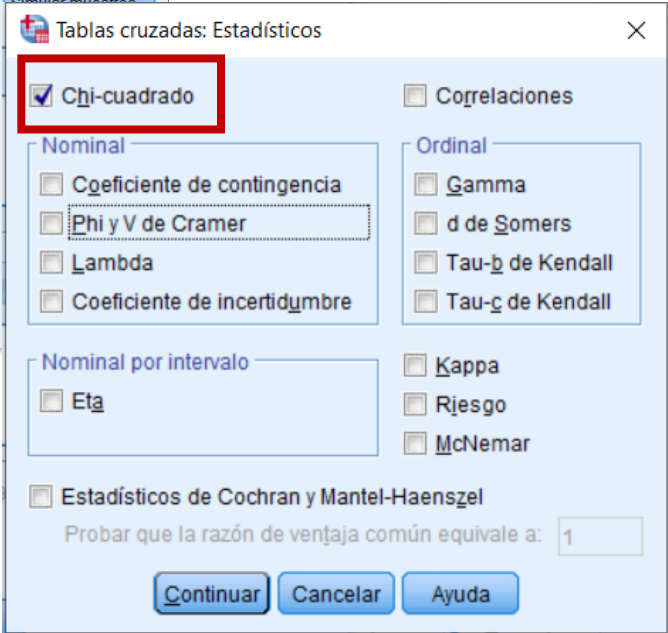
The 'Mostrar en las casillas' sub-dialog is shown in detail. It includes sections for 'Recuentos' (Observed, Expected, Hide small counts), 'Prueba z' (Compare proportions of column, Correct p-values), 'Residuos' (Non-standardized, Standardized, Standardized corrected), and 'Ponderaciones no enteras' (Round counts, Truncate counts, No corrections). The 'Porcentajes' section is highlighted with a red box, showing 'Ella', 'Columna', and 'Total' selected.

Prueba de independencia estadística χ^2

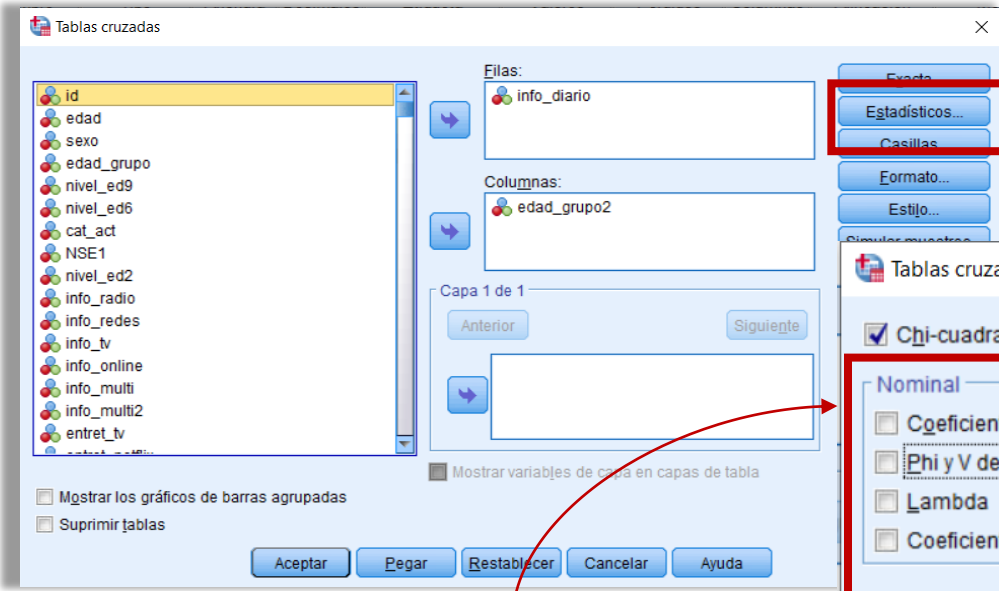


Botón > Estadísticos

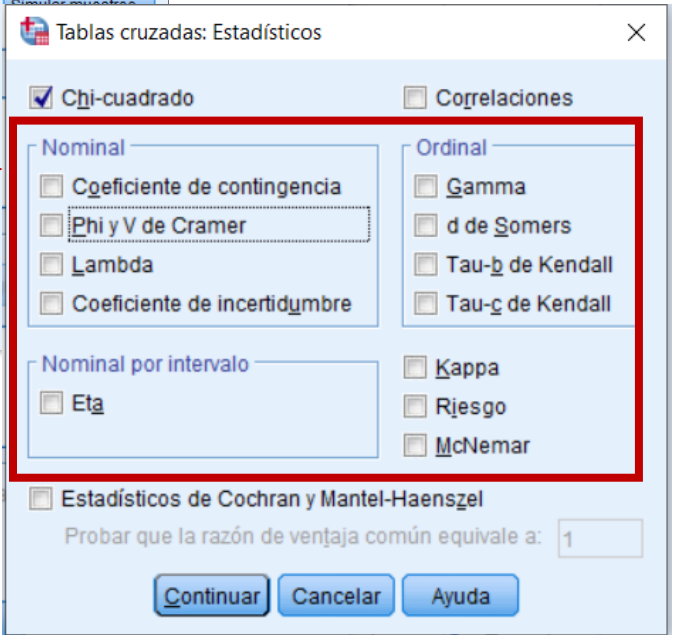
Seleccionamos la prueba



Coeficientes de asociación

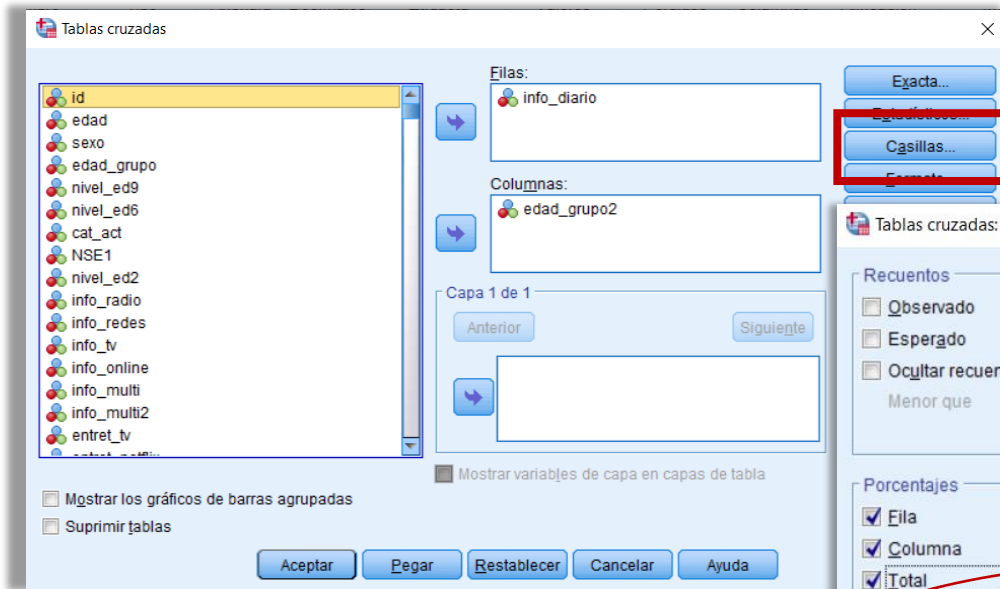


Botón > Estadísticos

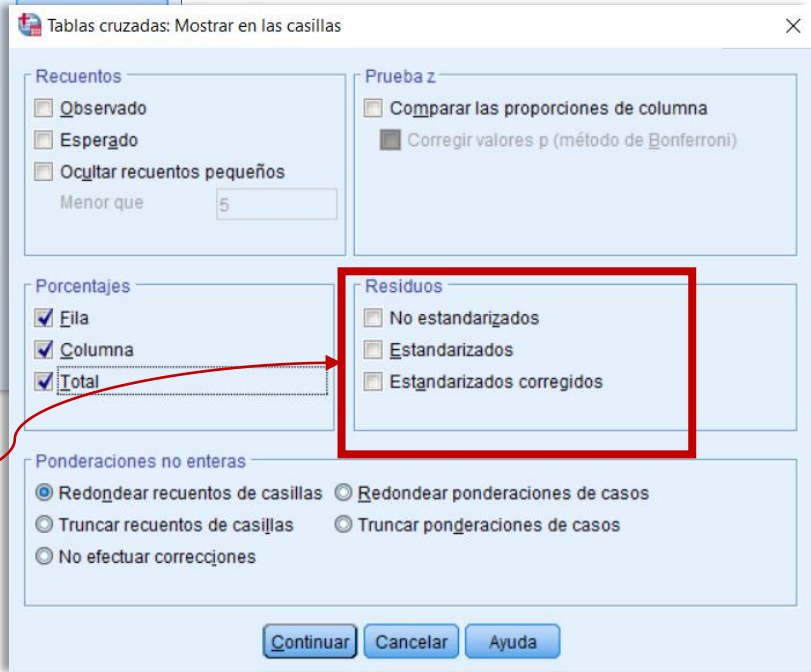


Seleccionamos el coeficiente

Análisis de residuos tipificados



Botón > Casillas



Seleccionamos los residuos

Aplicación

Planteo del problema:

La brecha laboral de género "se refiere a la relación entre las tasas de participación económica de varones y mujeres, subrayando las desigualdades existentes en el mercado laboral. Muestra la distancia que existe para obtener la igualdad de oportunidades en el mundo del trabajo" (Mazzeo & Bocchicchio, 2017: 3). La brecha laboral de género tendría una de sus expresiones en la desigual tasa de participación de las mujeres en el mercado de trabajo con respecto a los varones.

Hipótesis:

Existe una brecha laboral de género que se expresa en la participación laboral, de modo que las mujeres tendrán una menor probabilidad de participar en el mercado de trabajo que los varones.

Variables

Variable independiente: género (asume valores varón/mujer) → Nominal

Variable dependiente: condición de actividad (activo/inactivo) → Nominal

Aplicación

Consecuencia observacional de la hipótesis

Esquematizamos qué necesitaríamos observar en caso de que nuestra hipótesis se comprobase y qué resultados invalidarían nuestra hipótesis.

	Varón	Mujer
Activo	P	<i>p</i>
Inactivo		

Comprobaría la hipótesis ($p < P$)

	Varón	Mujer
Activo	P	P
Inactivo		

Refutaría la hipótesis ($P = P$)

Tipo de hipótesis

Evaluamos qué tipo de hipótesis tendríamos en función de la relación bivariada.

	Varón	Mujer
Activo		
Inactivo		

Diagonal

	Varón	Mujer
Activo		
Inactivo		

Rinconal

Aplicación: distintos %

Tabla cruzada activo Condición de actividad*sexo Sexo

			sexo Sexo		Total
			1,00 Varón	2,00 Mujer	
activo Condición de actividad	1,00 Activo/a	Recuento	10035	8080	18115
		% dentro de activo Condición de actividad	55,4%	44,6%	100,0%
	2,00 Inactivo/a	Recuento	1961	4989	6950
		% dentro de activo Condición de actividad	28,2%	71,8%	100,0%
Total		Recuento	11996	13069	25065
		% dentro de activo Condición de actividad	47,9%	52,1%	100,0%

% de columna

% de fila

% de total

Tabla cruzada activo Condición de actividad*sexo Sexo

			sexo Sexo		Total
			1,00 Varón	2,00 Mujer	
activo Condición de actividad	1,00 Activo/a	Recuento	10035	8080	18115
		% dentro de sexo Sexo	83,7%	61,8%	72,3%
	2,00 Inactivo/a	Recuento	1961	4989	6950
		% dentro de sexo Sexo	16,3%	38,2%	27,7%
Total		Recuento	11996	13069	25065
		% dentro de sexo Sexo	100,0%	100,0%	100,0%

Tabla cruzada activo Condición de actividad*sexo Sexo

			sexo Sexo		Total
			1,00 Varón	2,00 Mujer	
activo Condición de actividad	1,00 Activo/a	Recuento	10035	8080	18115
		% del total	40,0%	32,2%	72,3%
	2,00 Inactivo/a	Recuento	1961	4989	6950
		% del total	7,8%	19,9%	27,7%
Total		Recuento	11996	13069	25065
		% del total	47,9%	52,1%	100,0%

Aplicación: análisis de asociación

Resumen de procesamiento de casos

	Casos					
	Válido		Perdido		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
activo Condición de actividad * sexo Sexo	25065	100,0%	,106	0,0%	25065,106	100,0%

Tabla cruzada activo Condición de actividad*sexo Sexo

			sexo Sexo		Total
			1,00 Varón	2,00 Mujer	
activo Condición de actividad	1,00 Activo/a	Recuento	10035	8080	18115
		% dentro de sexo Sexo	83,7%	61,8%	72,3%
	2,00 Inactivo/a	Recuento	1961	4989	6950
		% dentro de sexo Sexo	16,3%	38,2%	27,7%
Total		Recuento	11996	13069	25065
		% dentro de sexo Sexo	100,0%	100,0%	100,0%

Comparar % y sacar conclusiones

Aplicación: medidas de asociación global

Pruebas de chi-cuadrado

	Valor	df	Significación asintótica (bilateral)	Significación exacta (bilateral)	Significación exacta (unilateral)
Chi-cuadrado de Pearson	1487,028 ^a	1	,000		
Corrección de continuidad ^b	1485,939	1	,000		
Razón de verosimilitud	1529,972	1	,000		
Prueba exacta de Fisher				,000	,000
Asociación lineal por lineal	1486,968	1	,000		
N de casos válidos	25065				

Evaluamos la significancia de la asociación

Evaluamos los distintos coeficientes

Medidas simétricas

		Valor	Error estándar asintótico ^a	T aproximada ^b	Significación aproximada
Nominal por Nominal	Phi	,244			,000
	V de Cramer	,244			,000
Ordinal por ordinal	Gamma	,519	,011	40,206	,000
N de casos válidos		25065			

Medidas direccionales

			Valor	Error estándar asintótico ^a	T aproximada ^b	Significación aproximada
Nominal por Nominal	Lambda	Simétrico	,103	,007	14,587	,000
		activo Condición de actividad dependiente	,000	,000	.c	.c
		sexo Sexo dependiente	,163	,010	14,587	,000
	Tau Goodman y Kruskal	activo Condición de actividad dependiente	,059	,003		,000 ^d
		sexo Sexo dependiente	,059	,003		,000 ^d

Aplicación: medidas de asociación local

Tabla cruzada activo Condición de actividad*sexo Sexo

			sexo Sexo		Total
			1,00 Varón	2,00 Mujer	
activo Condición de actividad	1,00 Activo/a	Recuento	10035	8080	18115
		% dentro de sexo Sexo	83,7%	61,8%	72,3%
		Residuo corregido	38,6	-38,6	
	2,00 Inactivo/a	Recuento	1961	4989	6950
		% dentro de sexo Sexo	16,3%	38,2%	27,7%
		Residuo corregido	-38,6	38,6	
Total	Recuento	11996	13069	25065	
	% dentro de sexo Sexo	100,0%	100,0%	100,0%	

Residuos altamente significativos

$$OR = \frac{10035/1961}{8080/4989} = \frac{5,11}{1,61} = 3,17$$

Entre los varones, las chances de ser activos frente a no serlo son 3,2 veces las que se registran entre las mujeres.