

Análisis de Estructuras Espaciales Persistentes. Desempleo Departamental en Argentina.

Marcos Herrera¹

¹CONICET

IELDE - Universidad Nacional de Salta (Argentina)

Seminario de Investigación N^o 27

8 de Agosto del 2013 (Salta, Argentina)



Esquema de Presentación

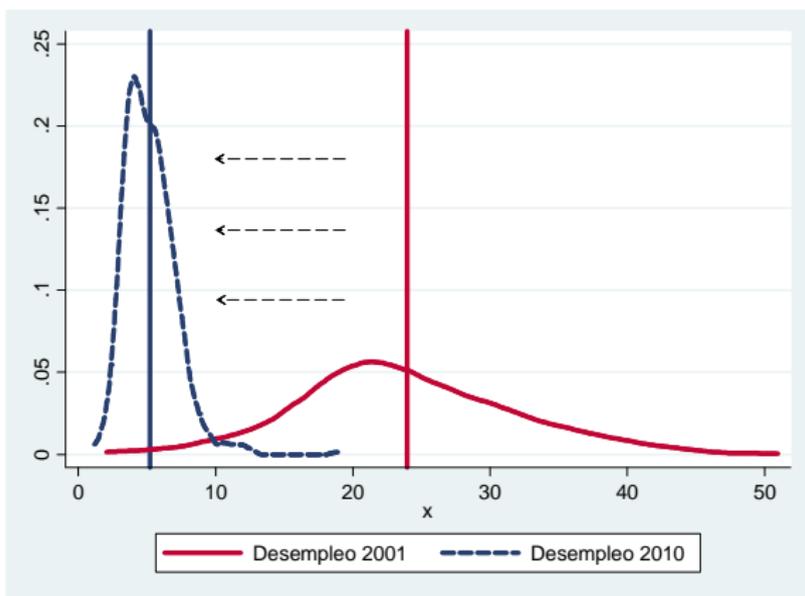
- 1 Motivación
- 2 Tests Espaciales Globales
 - Test I de Moran
 - Test C de Geary
 - Test Global de Getis y Ord
- 3 Tests Espaciales Locales
 - I de Moran Local
 - Test G^* de Getis y Ord
- 4 Análisis Simbólico
- 5 Análisis Desempleo
 - Tests Globales
 - I de Moran Local
 - Test G^* Getis-Ord
 - Test de Análisis Simbólico
- 6 Conclusiones

Esquema de Presentación

- 1 Motivación
- 2 Tests Espaciales Globales
 - Test I de Moran
 - Test C de Geary
 - Test Global de Getis y Ord
- 3 Tests Espaciales Locales
 - I de Moran Local
 - Test G^* de Getis y Ord
- 4 Análisis Simbólico
- 5 Análisis Desempleo
 - Tests Globales
 - I de Moran Local
 - Test G^* Getis-Ord
 - Test de Análisis Simbólico
- 6 Conclusiones

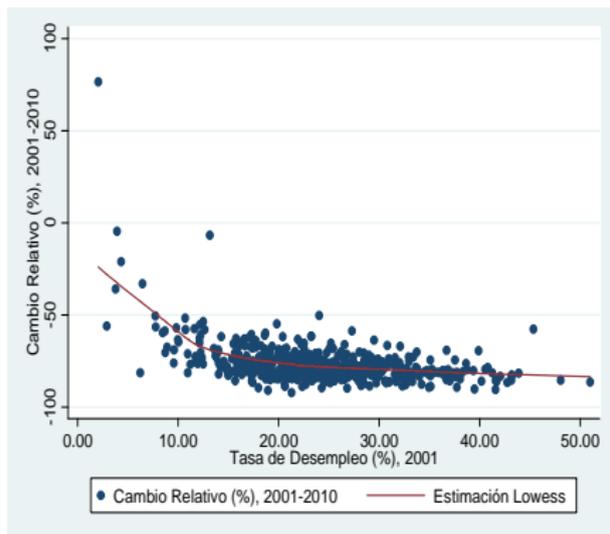
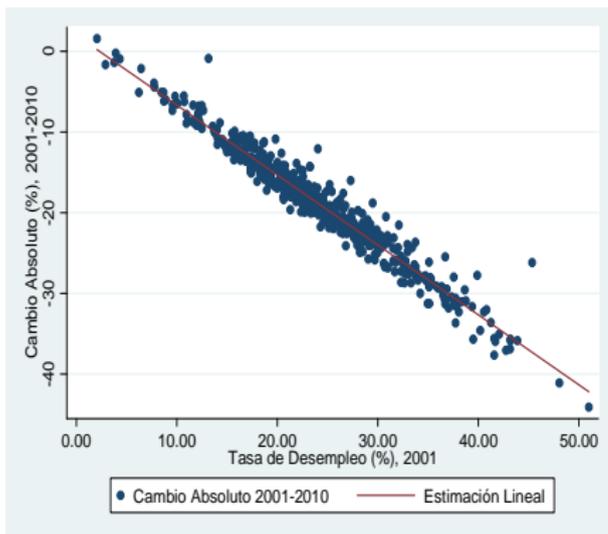
Evolución del Desempleo Departamental

Distribución de la Tasa de Desempleo. Censo 2001 y 2010.

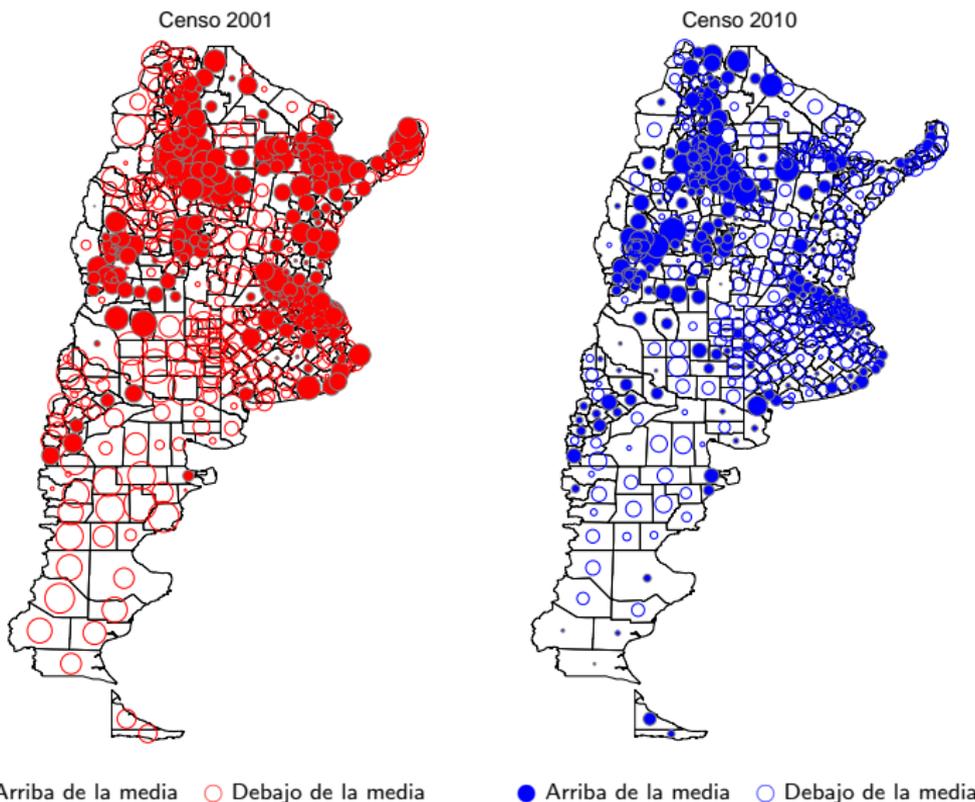


Las líneas verticales representan el valor promedio:
23.9 % (Censo 2001) y 5.2 % (Censo 2010).

Cambio Absoluto y Relativo del Desempleo



Distribución Espacial del Desempleo



Resumen Desempleo Departamental entre Censos

- 1 Marcado descenso del desempleo: de un valor promedio de 23.9 % en el 2001 a un valor promedio de 5.2 % en el 2010.
- 2 Los departamentos con mayor desempleo son los que experimentaron una mayor disminución absoluta.
- 3 La tasa de cambio relativa muestra que la mayor cantidad de los departamentos redujo su desempleo en mas del 50 % a lo largo de la década.
- 4 Sin embargo, la distribución espacial del desempleo muestra una estructura similar entre censos:
 - 1 Altos valores respecto a la media se sitúan en la misma zona geográfica en ambos censos.
 - 2 Bajos valores respecto a la media se encuentran en las mismas zonas en ambos censos.

El objetivo del trabajo es analizar mediante tests espaciales si existe evidencia significativa de persistencia espacial entre censos.

Esquema de Presentación

- 1 Motivación
- 2 Tests Espaciales Globales
 - Test I de Moran
 - Test C de Geary
 - Test Global de Getis y Ord
- 3 Tests Espaciales Locales
 - I de Moran Local
 - Test G^* de Getis y Ord
- 4 Análisis Simbólico
- 5 Análisis Desempleo
 - Tests Globales
 - I de Moran Local
 - Test G^* Getis-Ord
 - Test de Análisis Simbólico
- 6 Conclusiones

Test I de Moran

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad i \neq j \quad (1)$$

donde n es el número de unidades espaciales indexadas por i y j ; x es la variable de interés, \bar{x} es la media; y w_{ij} es un elemento de la matriz de pesos espaciales.

El valor esperado del test es igual a $-1/(n-1)$.

La hipótesis nula es no autocorrelación espacial y el estadístico tiene una distribución normal que suele aproximarse por permutación.

Test C de Geary

$$C = \frac{(n-1) \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - x_j)^2}{2 \left(\sum_{i=1}^n \sum_{j=1}^n w_{ij} \right) \sum_{i=1}^n (x_i - \bar{x})^2}, \quad i \neq j \quad (2)$$

donde la notación es la misma que para el Test I de Moran. El test asume un rango entre 0 y 2. El valor esperado es 1. Valores inferiores a 1 (mayores a 1) indican autocorrelación espacial negativa (positiva). La hipótesis nula es la misma que el I de Moran y se distribuye asintóticamente normal.

Test de Getis - Ord Global

Finalmente, se presenta la versión global para el test de Getis y Ord:

$$G = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} x_i x_j}{\sum_{i=1}^n \sum_{j=1}^n x_i x_j}, \quad i \neq j \quad (3)$$

donde se continúa con la misma notación.

La hipótesis nula del test es no agrupamiento espacial general y asintóticamente el estadístico se distribuye normal.

Esquema de Presentación

- 1 Motivación
- 2 Tests Espaciales Globales
 - Test I de Moran
 - Test C de Geary
 - Test Global de Getis y Ord
- 3 Tests Espaciales Locales
 - I de Moran Local
 - Test G^* de Getis y Ord
- 4 Análisis Simbólico
- 5 Análisis Desempleo
 - Tests Globales
 - I de Moran Local
 - Test G^* Getis-Ord
 - Test de Análisis Simbólico
- 6 Conclusiones

Moran Local

$$I_i(d) = \frac{(x_i - \bar{x})}{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \sum_{j=1, j \neq i}^n w_{ij}(d) (x_j - \bar{x}), \quad (4)$$

donde d representa al radio de distancia elegido.

Este test permite detectar grupos de observaciones que pueden clasificarse en Alta-Alto (H-H), Bajo-Bajo (L-L), Bajo-Alto (L-H) y Alto-Bajo (H-L).

Estas áreas pueden fácilmente visualizarse en el gráfico de dispersión de Moran.

Test Local G^* de Getis - Ord

$$G_i^*(d) = \frac{\sum_{j=1}^n w_{ij}(d) x_j}{\sum_{j=1}^n x_j}, \quad (5)$$

donde para $w_{ij} = 1$, tal que el valor del estadístico considera el valor de la variable en el punto i .

Este test detecta grupos observaciones de bajos y altos valores de la variable en un entorno local (definido por el parámetro d).

Esquema de Presentación

- 1 Motivación
- 2 Tests Espaciales Globales
 - Test I de Moran
 - Test C de Geary
 - Test Global de Getis y Ord
- 3 Tests Espaciales Locales
 - I de Moran Local
 - Test G^* de Getis y Ord
- 4 **Análisis Simbólico**
- 5 Análisis Desempleo
 - Tests Globales
 - I de Moran Local
 - Test G^* Getis-Ord
 - Test de Análisis Simbólico
- 6 Conclusiones

Introducción al Análisis Simbólico

- *Análisis simbólico* se basa en la transformación de una serie dentro de una secuencia de símbolos que capturan información que no puede ser observada directamente.
 - Dado un proceso $\{y_s\}$, definimos una función $f : \{y_s\} \rightarrow \Gamma_n$, donde f es la *función simbolizadora* y Γ_n es el conjunto de *símbolos* (σ 's).
- Conceptos Clave:
 - m es la *dimensión de encaje*: nos brinda la longitud del vecindario.
 - $m - \text{vecindario}$: es el conjunto formado por la información de cada observación y sus $(m - 1)$ vecinos. *El $m - \text{vecindario}$ es la información que usamos para construir el símbolo σ .*
- Función Simbolizadora Utilizada: Símbolo (σ_{s_j}) captura el número de vecinos que tienen igual valor al de la localización s_j .

Test Ψ_2 para comparar mapas

Entropía es una medida de incertidumbre asociada con la distribución de una variable aleatoria. Para una variable aleatoria discreta, la *entropía simbólica* de $\{y_s\}$:

$$h_y(m) = - \sum_{\sigma \in \Gamma} p_\sigma \ln(p_\sigma).$$

donde p_σ es la probabilidad del símbolo σ .

La *entropía simbólica conjunta* es:

$$h_Z(m) = - \sum_{Y_{st}} \sum_{Y_{st+j}} p_{\sigma_{Y_t} \sigma_{Y_{t+j}}} \ln(p_{\sigma_{Y_t} \sigma_{Y_{t+j}}})$$

La expresión del estadístico es $\Psi_2 = h_{y_s(t)}(m) + h_{y_s(t+j)}(m) - h_Z(m)$ siendo la hipótesis nula:

$$H_0 : \{y_s\}_t \text{ and } \{y_s\}_{t+j} \text{ son independientes.} \quad (6)$$

donde $\{y_s\}$ es un proceso espacial en el periodo t y $t + j$.

Bajo H_0 , $\Psi_2 = 0$. El nivel de significancia es obtenido por bootstrap de bloques espaciales que preserva la dependencia espacial de la serie en cada periodo temporal.

Esquema de Presentación

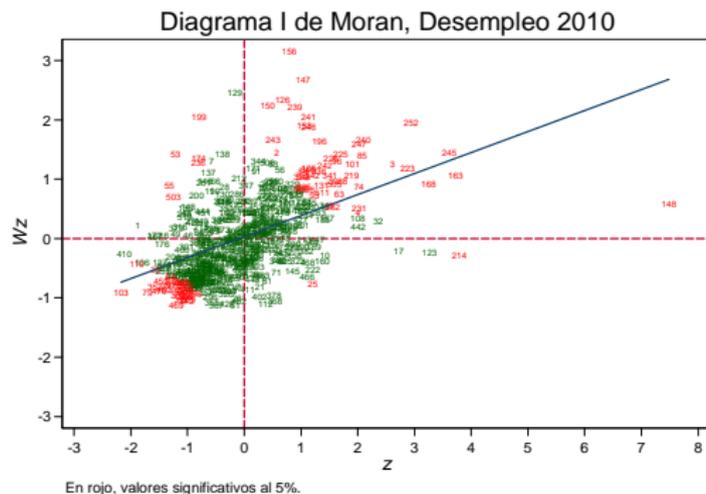
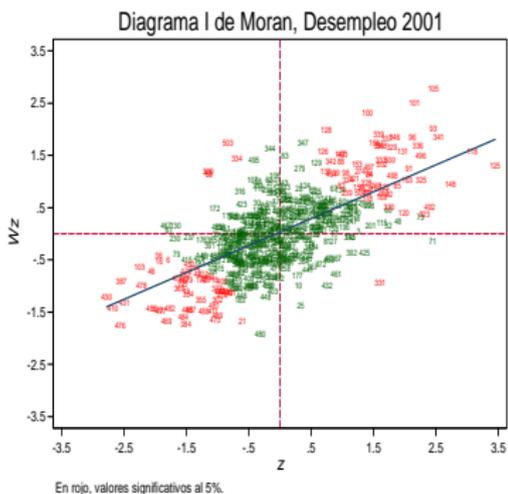
- 1 Motivación
- 2 Tests Espaciales Globales
 - Test I de Moran
 - Test C de Geary
 - Test Global de Getis y Ord
- 3 Tests Espaciales Locales
 - I de Moran Local
 - Test G^* de Getis y Ord
- 4 Análisis Simbólico
- 5 Análisis Desempleo
 - Tests Globales
 - I de Moran Local
 - Test G^* Getis-Ord
 - Test de Análisis Simbólico
- 6 Conclusiones

Tests de Dependencia Espacial Globales

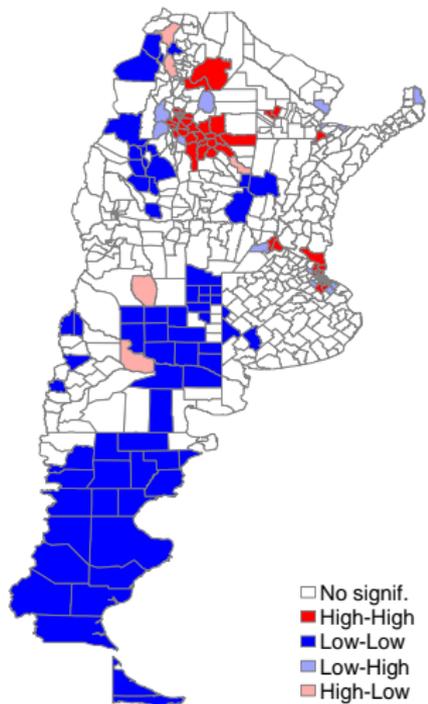
Variables	Desempleo, 2001		Desempleo, 2010	
Test	Valor	p-valor	Valor	p-valor
I de Moran	0.51	0.00	0.35	0.00
c de Geary	0.50	0.00	0.65	0.00
G de Getis-Ord	0.01	0.00	0.01	0.00

Se detecta estructura espacial en cada censo. Esto no significa que la estructura espacial sea la misma. Debemos utilizar otras herramientas para obtener una comparación entre censos

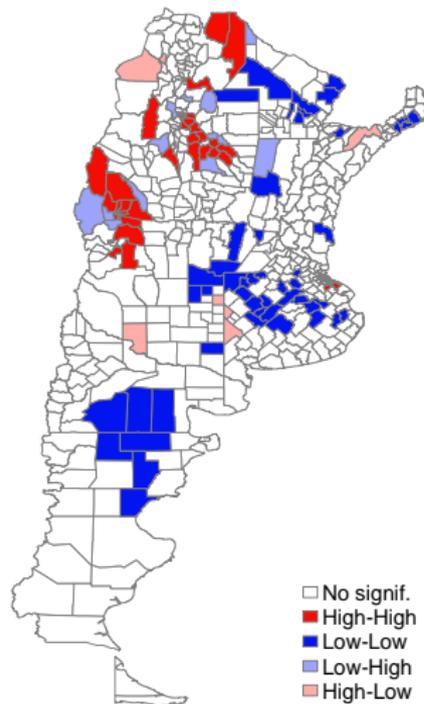
Diagrama I de Moran con Valores Significativos



Clusters I Moran Local



Clusters LISA, 2001



Clusters LISA, 2010

Tests de Independencia

<i>LISA</i> <i>Test</i>		Censo 2010					
		No sig.	H-H	L-L	L-H	H-L	Total
Censo 2001	No sig.	292	22	40	10	1	365
	H-H	31	26	0	2	0	59
	L-L	50	0	10	0	5	65
	L-H	14	1	0	1	0	16
	H-L	5	0	0	0	0	5
	Total	392	49	50	13	6	510

Dada la tabla de contingencia, la hipótesis nula de independencia de dos respuestas es $H_0 : \pi_{ij} = \pi_{i+}\pi_{+j}$

$$Pearson \chi^2 (16) = 130,5$$

$$p - valor = 0,00$$

$$LR \chi^2 (16) = 130,5$$

$$p - valor = 0,00$$

Tests de Independencia

<i>LISA</i> <i>Test</i>		Censo 2010					Total
		No sig.	H-H	L-L	L-H	H-L	
Censo 2001	No sig.	292	22	40	10	1	365
	H-H	31	26	0	2	0	59
	L-L	50	0	10	0	5	65
	L-H	14	1	0	1	0	16
	H-L	5	0	0	0	0	5
	Total	392	49	50	13	6	510

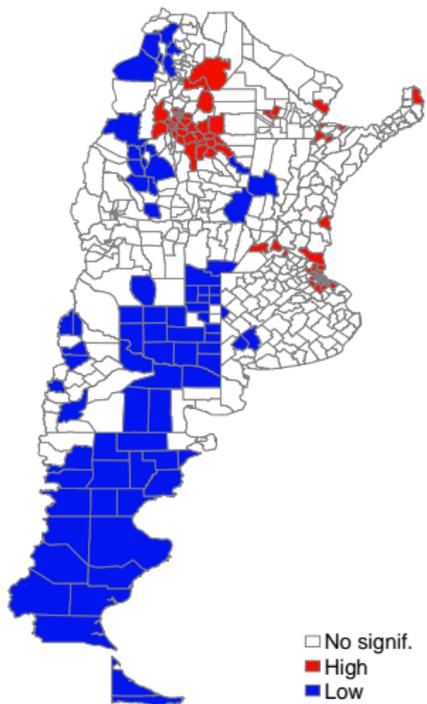
Dada la tabla de contingencia, la hipótesis nula de independencia de dos respuestas es $H_0 : \pi_{ij} = \pi_{i+}\pi_{+j}$

$$Pearson \chi^2 (16) = 130,5$$

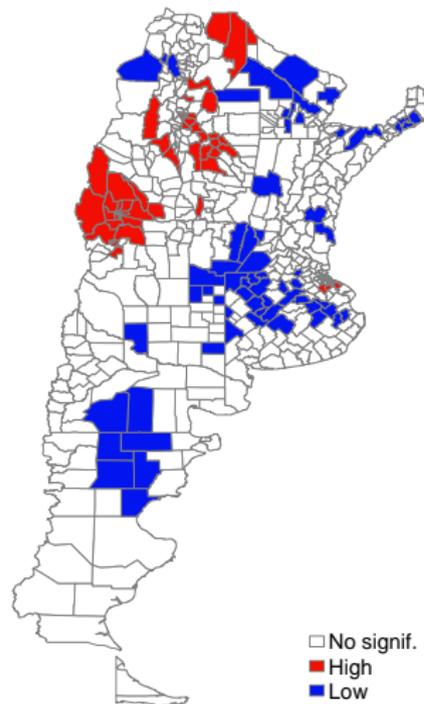
$$p - valor = 0,00$$

$$LR \chi^2 (16) = 130,5$$

$$p - valor = 0,00$$

Test G^* Getis-OrdClusters G^* Getis-Ord

Clusters Getis-Ord, 2001



Clusters Getis-Ord, 2010

Tests de Independencia

$Test G^*$ Getis-Ord		Censo 2010			
		No sig.	High	Low	Total
Censo 2001	No sig.	281	32	46	359
	High	47	29	0	76
	Low	57	1	17	75
	Total	385	62	63	510

Dada la tabla de contingencia, la hipótesis nula de independencia de dos respuestas es $H_0 : \pi_{ij} = \pi_{i+}\pi_{+j}$

$$Pearson \chi^2(4) = 70,8$$

$$p - valor = 0,00$$

$$LR \chi^2(4) = 68,5$$

$$p - valor = 0,00$$

Tests de Independencia

$Test G^*$ Getis-Ord		Censo 2010			
		No sig.	High	Low	Total
Censo 2001	No sig.	281	32	46	359
	High	47	29	0	76
	Low	57	1	17	75
	Total	385	62	63	510

Dada la tabla de contingencia, la hipótesis nula de independencia de dos respuestas es $H_0 : \pi_{ij} = \pi_{i+}\pi_{+j}$

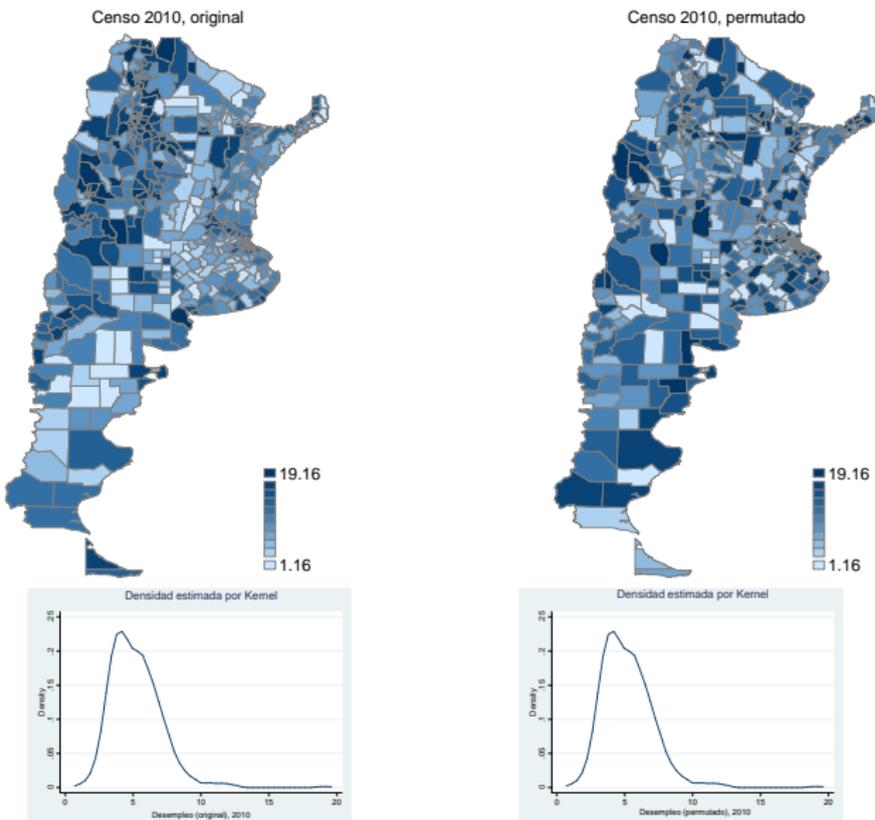
$$Pearson \chi^2(4) = 70,8$$

$$p - valor = 0,00$$

$$LR \chi^2(4) = 68,5$$

$$p - valor = 0,00$$

Distribución Original y Permutada, 2010



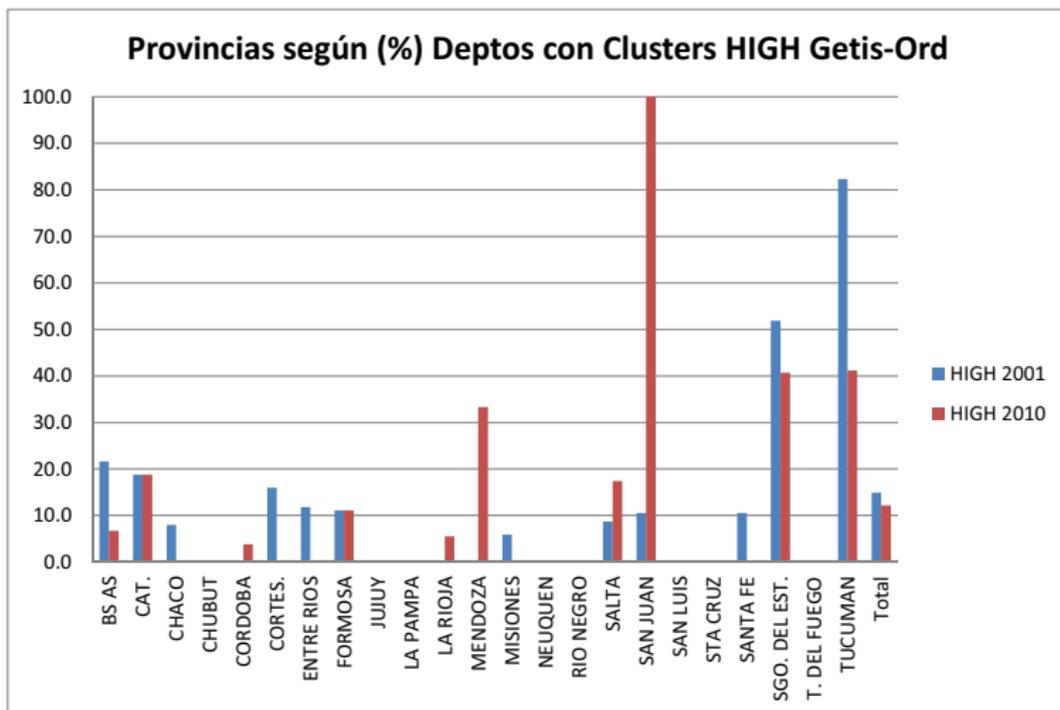
Tests de Independencia

Test de Independencia entre Mapas			
Test Ψ_2	Valor	p-valor	Conclusión
Desemp. 2010 (original) - Desemp. 2001	0,068	0,002	Mapas dependientes
Desemp. 2010 (permut.) - Desemp. 2001	0,011	0,835	Mapas independientes

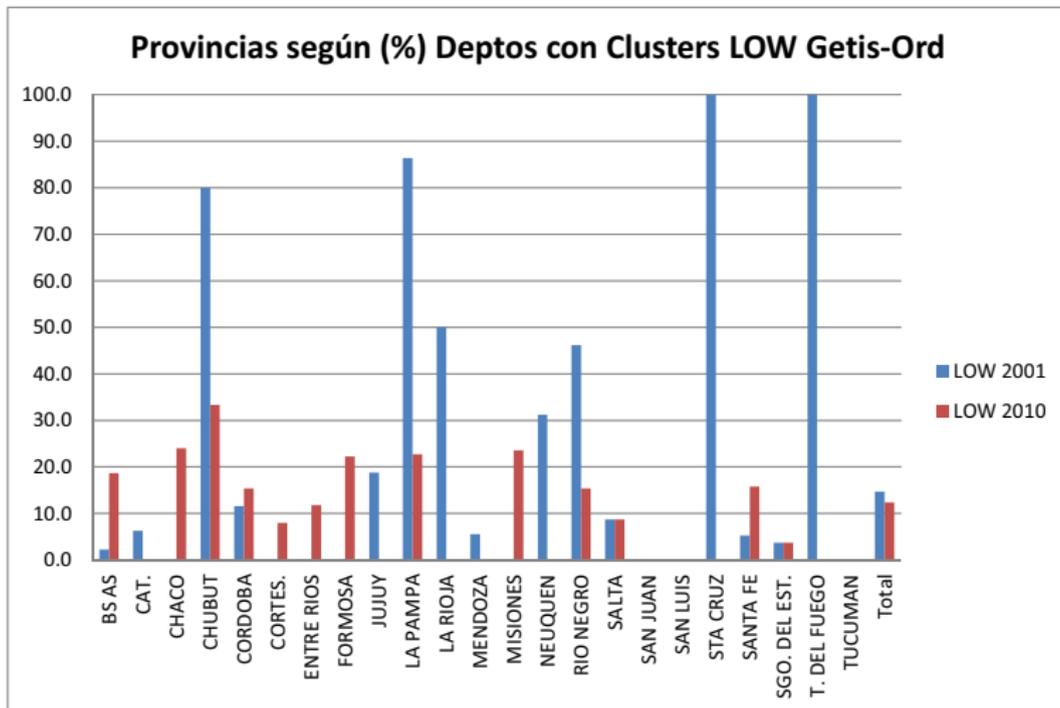
Nota: W-4nn, 10 bloques, 399 boots.

Este test detecta dependencia entre mapas, conservando la estructura espacial de cada uno de ellos. Sin embargo, es un test de dependencia global, no puede detectar clusters significativos.

¿Y qué decir respecto a las provincias?



¿Y qué decir respecto a las provincias?



Esquema de Presentación

- 1 Motivación
- 2 Tests Espaciales Globales
 - Test I de Moran
 - Test C de Geary
 - Test Global de Getis y Ord
- 3 Tests Espaciales Locales
 - I de Moran Local
 - Test G^* de Getis y Ord
- 4 Análisis Simbólico
- 5 Análisis Desempleo
 - Tests Globales
 - I de Moran Local
 - Test G^* Getis-Ord
 - Test de Análisis Simbólico
- 6 Conclusiones

Resumen

- A pesar de la marcada disminución del desempleo en prácticamente todos los departamentos entre los censos 2001 y 2010:
 - Se detecta dependencia espacial global en cada censo.
 - Los tests locales detectan clusters espaciales en cada censo.
- La dependencia espacial persiste entre ambos censos:
 - Las tablas de contingencia de los clusters espaciales rechaza la hipótesis nula de independencia: existen estructuras espaciales persistentes.
 - El test Ψ_2 , basado en análisis simbólico, confirma que ambos mapas son dependientes: la dependencia espacial persiste en el tiempo.
- Futuros desarrollos:
 - Corrección de los test de contingencia por dependencia entre categorías de clusters.
 - Descomposición del test de análisis simbólico para la detección local de estructuras espaciales persistentes en el tiempo.

Gracias por su Atención