

# Macrolocalizacion de regiones con K-means

## Índice

Carga de datos . . . . .	1
Normalización de la base de datos . . . . .	1
Número de clusters óptimo . . . . .	1
Cómputo de las K-means . . . . .	2
Mapa de las provincias Argentinas divididas por Cluster . . . . .	4
Flujos migratorios entre clusters . . . . .	4
Definición de migrante según cluster de origen . . . . .	5
Factores determinantes de expulsión de los migrantes . . . . .	6
Factores determinantes de atracción de los migrantes . . . . .	8
## Joining, by = c("CODUSU", "ANO4", "TRIMESTRE", "NRO_HOGAR", "REGION", "MAS_500", "AGLOMERADO", "POND")	
## Joining, by = "AGLOMERADO"	
## Joining, by = "SECTOR"	

## Carga de datos

Para comenzar, cargamos los datos de las variables sociales y economicas que definen a las distintas provincias de la Argentina.

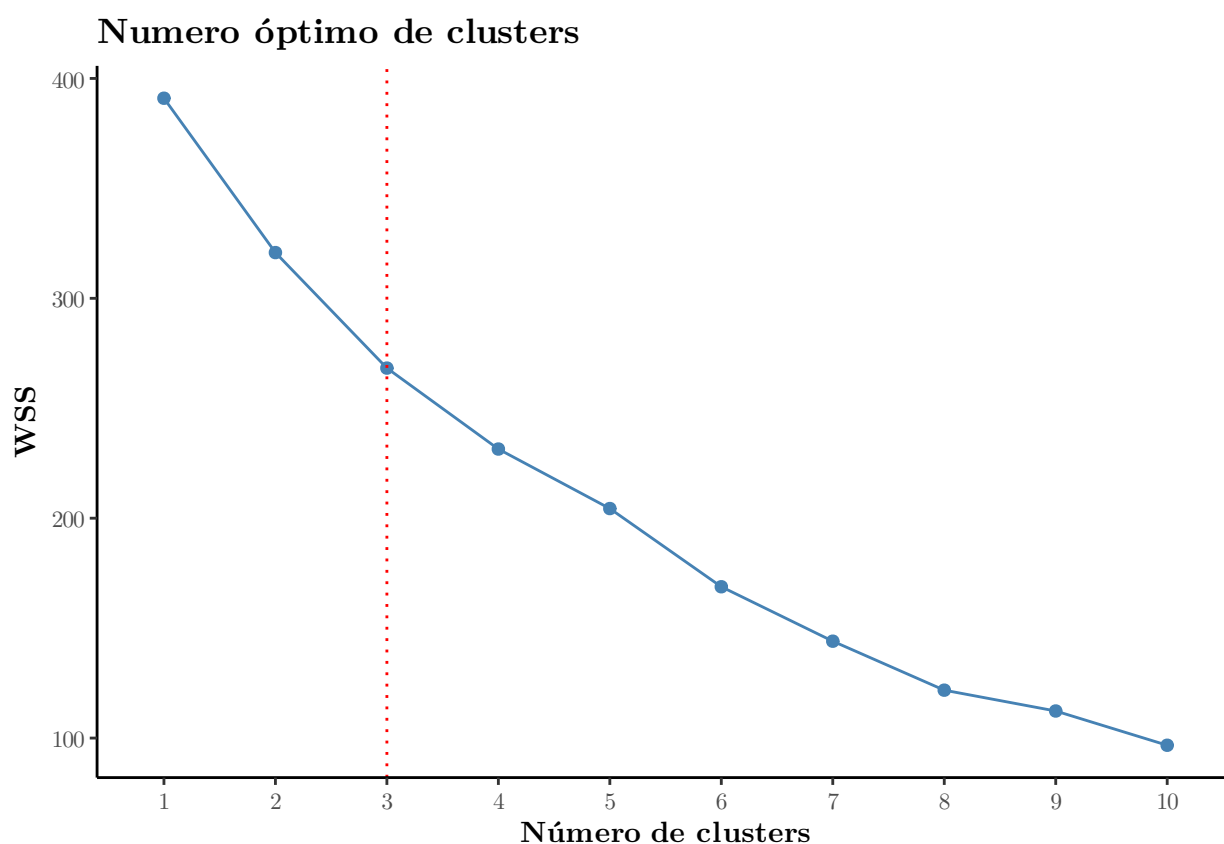
## Normalización de la base de datos

Ahora normalizamos la base de datos para lograr coherencia en la aplicación de K-means, debido a que trabajamos con distintas escalas de datos.

## Número de clusters óptimo

Verificamos el numero de clusters optimos a través de el Within-cluster sum of square, el cual es un indicador que mide la suma de las distancias entre las variables dentro de los cluster y sus centroides, la idea es minimizar esta discrepancia dentro de cada grupo, teniendo en cuenta el trade-off que implica cuando este es mínimo, el cual es el caso en que el numero de clusters es igual al número de variables a clusterizar, en donde si tenemos  $j$  observaciones, el numero de clusters sería tal que  $j=k$ , y no se estaría dando ninguna información relevante a los efectos de poder resumir características comunes entre los grupos.

$$WCSS = \sum_{i=1}^{N_C} \sum_{\mathbf{x} \in C_i} distancia(\mathbf{x}, \bar{\mathbf{x}}_{C_i})$$



Obtenemos que el numero óptimo de clusters se encuentra en torno al **número 3**, debido a que la contribución marginal de aumentar el numero de clusters a 4, no aportaría una reducción muy elevada al **WSS**, y seguiríamos perdiendo generalidad en la clusterización de los grupos sin una ganancia significativa de similitud dentro de los grupos.

## Cómputo de las K-means

Se realiza un seteo de una semilla para el cálculo de los centroides iniciales del algoritmo de K-means y se procederá a realizar el cálculo de los clusters a través de 1000 iteraciones, eligiendo la que arroje un menor numero de WSS.

Podemos ver la media de los parámetros para cada uno de los clusters

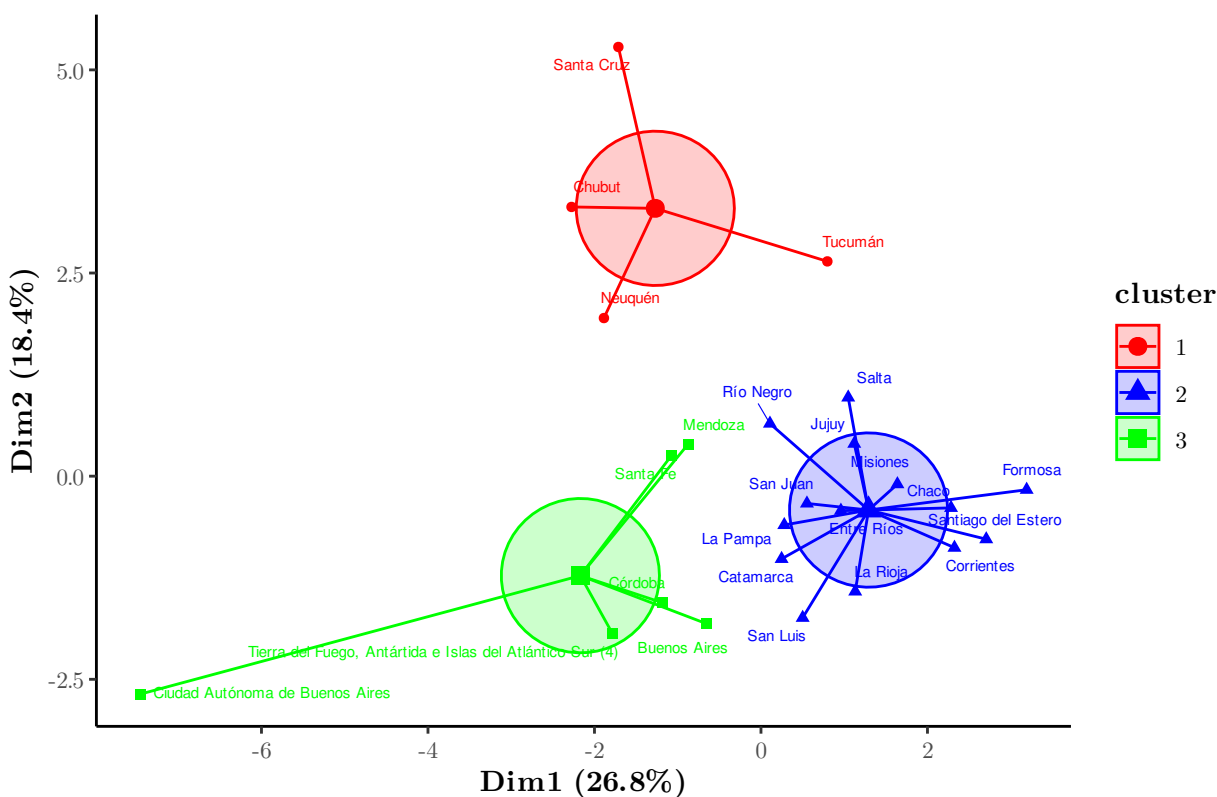
Cuadro 1: Resumen de indicadores por cluster

Indicadores	Cluster 1	Cluster 2	Cluster 3
Habitantes (2010)	530,187.00	677,181.00	3,042,344.00
Tasa promocion efectiva secundaria (2017)	76.25	78.60	82.05
Mortalidad infantil promedio 2016-2019	8.20	9.50	8.25
Homicidios dolosos C/ 100.000 hab. promedio (2016-2019)	7.25	3.75	5.30
Muertes en Accidentes Viales C/ 100.000 hab. promedio (2016-2019)	12.40	14.05	7.50
Robos (excluye los agravados) C/ 100.000 hab. Promedio (2016-2019)	973.40	654.85	1,566.95
Robos agravados C/ 100.000 hab. Promedio (2016-2019)	27.80	1.70	9.50
Violaciones C/ 100.000 hab. Promedio (2016-2019)	15.45	11.15	8.75
Exportaciones per-cápita en USD promedio (2016-2019)	2,857.20	794.20	1,420.55
Demanda de MWH energía electrica per cápita (2016)	3.80	2.20	3.35
Pobreza Promedio (2017-2019)	26.05	32.55	29.35

Indicadores	Cluster 1	Cluster 2	Cluster 3
Tasa actividad Promedio (2017-2019)	44.55	42.50	46.00
Cantidad empresas C/ 100.000 hab. (2016-2017)	1,569.62	846.96	1,771.95
Remuneracion real de trabajadores registrados del sector privado (2016-2019)	29,995.27	14,997.52	17,149.25
Porcentaje de empleados en Agricultura, ganadería, pesca y actividades extractivas (2016-2019)	0.21	0.13	0.05
Porcentaje de empleados en Comercio, servicios, electricidad, gas, agua y construccion (2016-2019)	0.67	0.69	0.71
Porcentaje de empleados en Industria (2016-2019)	0.10	0.15	0.22


agregamos a la base de datos original la denominación de cada cluster

### Plot de clusters resultantes de la macrolocalización



Se puede ver claramente la diferencia en la similitud de los tres grupos, siendo solamente la Capital Federal la única que presenta mayor disimilitud con respecto a su cluster, lo cual indica que esta capital lleva una dinámica socio-económica muy peculiar con respecto al promedio de las provincias argentinas, inclusive considerablemente distinta que las provincias con la cual ostenta mayor similitud.

A map of Argentina with its provinces colored to represent political affiliations. The colors are Blue, Green, and Red. The map shows a distribution where Blue is the most prevalent color, covering a large portion of the northern and central regions. Green is concentrated in the central and southern regions, and Red is found in the southern and western regions.



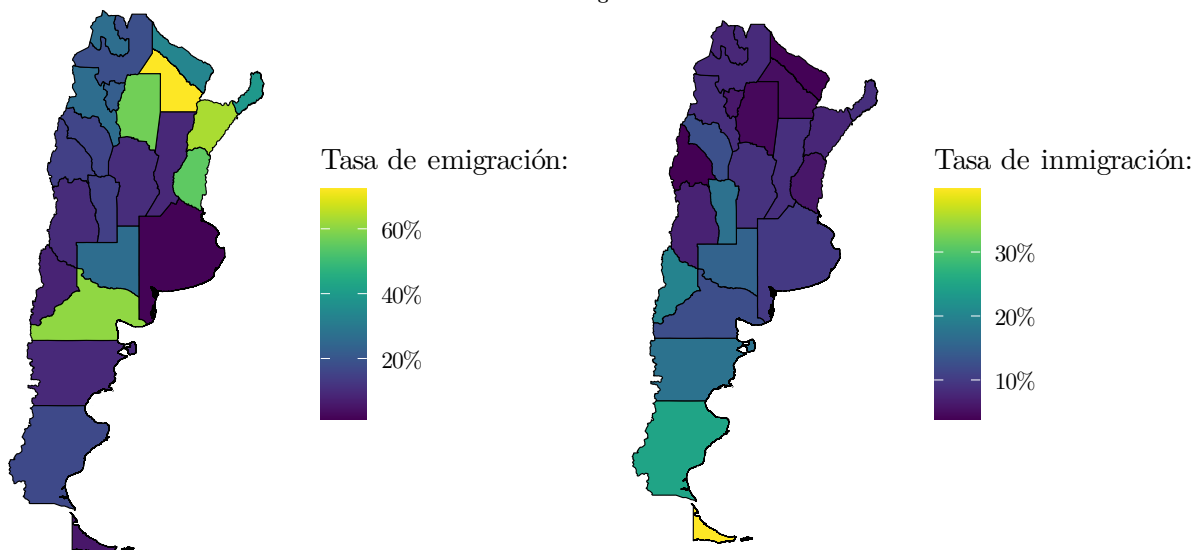
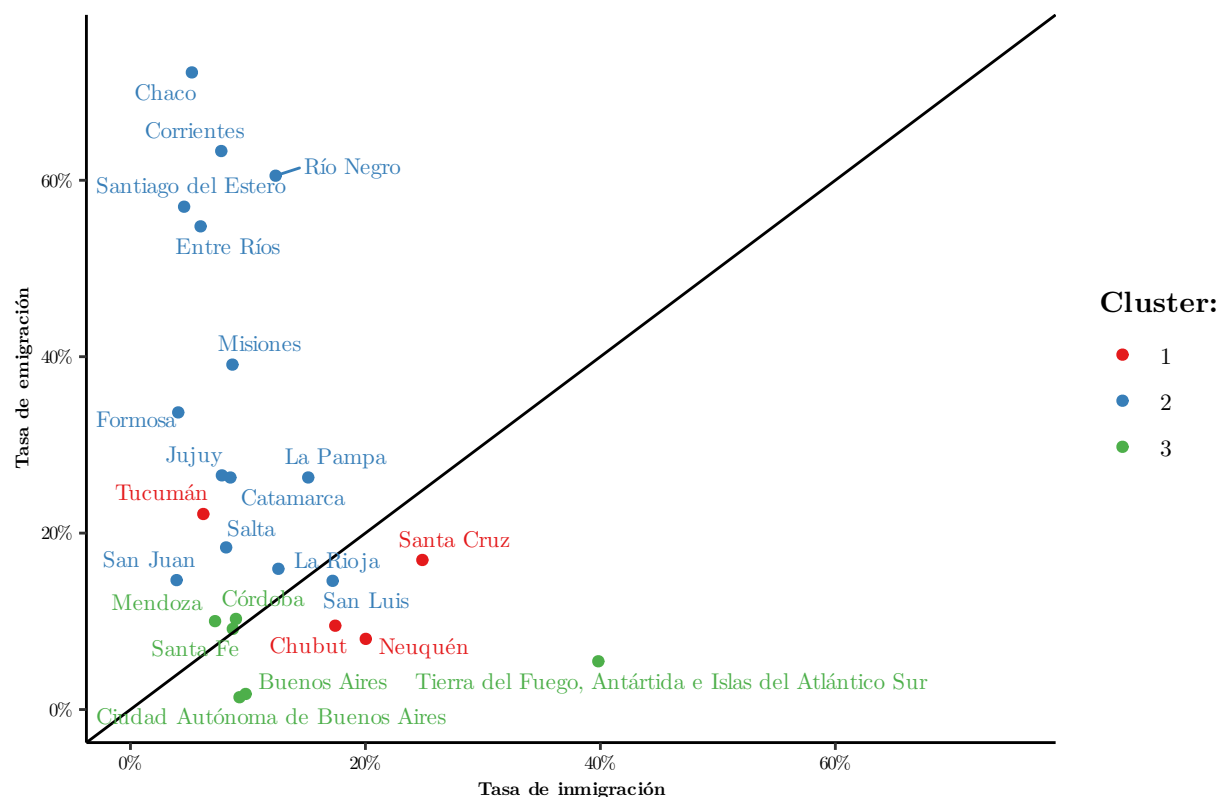
Dentro de los cluster, existen provincias de las cuales salen gran cantidad de inmigrantes, al igual que existen provincias que son receptoras de estos mismos, en primer lugar caracterizaremos cual es el cluster con mayor nivel de *expulsion* de inmigrantes interprovinciales.

Cluster	Porcentaje
1	10.41 %
2	61.67 %
3	27.92 %

Cluster	Porcentaje
1	7.85 %
2	12.83 %
3	79.32 %

4

## Tasa de inmigración y emigración por provincia



## Definición de migrante según cluster de origen

En primer lugar se analizará si existen microdeterminantes dentro de las características de los migrantes que provoquen que ciertas personas tengan una mayor propensión a migrar desde ciertas localizaciones, es decir, se analizarán los factores de expulsión de los distintos cluster; por otro lado, se realizará un análisis diferenciando el destino de los distintos migrantes, con el fin de establecer los determinantes de atracción de las distintas localizaciones.

## Factores determinantes de expulsión de los migrantes

En el universo de habitantes de una determinada región/cluster, pueden existir características determinadas que diferencien de manera significativa entre las personas que nacieron y decidieron quedarse en la misma región, de las que si bien nacieron en la misma región, optaron por migrar hacia otra. A continuación se realizará un análisis diferenciando por cluster de características socioeconómicas de los migrantes y no migrantes que nacieron en un mismo cluster, con el fin de determinar si existen diferencias significativas entre ellos.

Los factores determinantes de expulsión de los migrantes varían dependiendo del cluster que se esté tomando a consideración, el sexo no es un factor determinante de expulsión para los cluster 1 y 2, sin embargo, si lo es para el cluster 3, en el cual los hombres que nacieron en el mismo tienen una mayor probabilidad de migrar que las mujeres; la edad, por otro lado, tiene una relación positiva con la probabilidad de migrar para los tres clusters, lo que indica que la probabilidad de migrar aumenta conforme la persona es mayor, particularmente para el cluster 3 existe una relación cuadrática significativa, en donde cada año que pasa esta probabilidad crece a tasa decreciente, lo que genera que cada año que aumente la edad aportará cada vez menos en el aumento de la probabilidad de ser migrante.

La incidencia en la pobreza es un factor determinante significativo en la migración en los tres clusters, para los cluster 1 y 2 las personas pobres que nacieron en los mismos tienen una mayor probabilidad de abandonar sus provincias de origen, mientras que la incidencia en la pobreza para los oriundos del cluster 3 tiene un efecto totalmente inverso, reduciendo la probabilidad de abandonar su provincia natal.

La recepción de subsidios o ayudas por parte del gobierno, instituciones, iglesias, etc, tiene un impacto significativo en la probabilidad de migración en los tres clusters, pero nuevamente se puede ver un patrón diferenciado entre los cluster 1 y 2, en donde la recepción de un subsidio disminuye la probabilidad de ser migrante, mientras que en el cluster 3 el efecto es contrario, en donde la recepción de esta transferencia económica aumenta la probabilidad de abandonar la provincia de origen.

En cuanto a la condición de ocupación y la calificación del trabajo, los cluster 1 y 3 tienen efectos similares en cuanto a este factor determinante de la migración, en donde la probabilidad de migrar se reduce en presencia de un trabajo de baja calificación en relación a una persona que está desocupada o inactiva (en adelante grupo base), este efecto se intensifica y se generaliza en los tres clusters si la condición de ocupación de la persona es en un trabajo de alta calificación, en donde la probabilidad de migrar se reduce en comparación con el grupo base, lo cual indica que una persona con un trabajo de calificación alta tiene menos probabilidad de ser migrante de su provincia de origen para los tres clusters.

Otro factor relevante a la hora de analizar los factores de expulsión de los clusters es la organización familiar, en este caso, en el cluster 1 y 2 ser soltero con hijos disminuye la probabilidad de abandonar la provincia de origen en comparación con una persona soltera sin hijos (en adelante grupo base), mientras que en el cluster 3 el efecto es inverso, donde ser soltero con hijos aumenta la probabilidad de migrar, en relación con el grupo base. La similitud entre el cluster 1 y 2 se vuelve a repetir si consideramos el aumento en la probabilidad de ser migrante en el caso de ser estar en pareja (casado o unido) SIN hijos, en comparación con el grupo base, en donde la condición de estar en pareja reforzaría la decisión de migración, lo cual queda revertido en el caso de las personas que están casadas pero con hijos, el cual para todos los clusters es un factor determinante negativo que reduce la probabilidad de migrar en relación al grupo base y es estadísticamente significativo.

La propiedad de la vivienda es otro factor que tiene efectos diferenciados en la decisión de migración dependiendo del cluster que se esté analizando, para los cluster 1 y 2, el hecho de ser propietario/a de la vivienda reduce la probabilidad de ser un migrante con origen en estas provincias en relación con las personas que no son propietarias de las viviendas (en adelante grupo base), sin embargo, para el cluster 3 este efecto es inverso, el cual indica que la propiedad de la vivienda aumenta la probabilidad de haber abandonado su provincia de origen en relación con el grupo base.

Por último se analiza el efecto que tiene la educación en la expulsión de la provincia de origen, para el cluster 1 una persona con nivel educativo medio (secundaria completa) tiene menor probabilidad de haber abandonado la provincia de origen en relación con una persona con un nivel educativo bajo (en adelante grupo base), sin embargo, para los cluster 2 y 3 esta probabilidad es inversa, en donde una persona con nivel educativo

Cuadro 4:

	<i>Variable dependiente:</i>		
	Inmigrante		
	(Cluster 1)	(Cluster 2)	(Cluster 3)
Hombre	0.008 (0.037)	0.031 (0.020)	0.121*** (0.027)
Pobre	0.449*** (0.035)	0.058*** (0.020)	-0.476*** (0.026)
Subsidio	-0.125** (0.052)	-0.591*** (0.033)	0.325*** (0.040)
Ocupado/a baja calif.	-0.092** (0.041)	0.004 (0.023)	-0.190*** (0.031)
Ocupado/a alta calif.	-0.248*** (0.057)	-0.030 (0.032)	-0.177*** (0.036)
Soltero/a con Hijo/a	-0.311*** (0.048)	-0.388*** (0.027)	0.092** (0.036)
Casado/a sin Hijo/a	0.193*** (0.052)	0.109*** (0.027)	0.002 (0.035)
Casado/a con Hijo/a	-0.249*** (0.045)	-0.254*** (0.024)	-0.116*** (0.032)
Propietario/a de Vivienda	-0.214*** (0.035)	-0.545*** (0.019)	0.087*** (0.025)
Edad	0.043*** (0.006)	-0.007** (0.003)	0.064*** (0.005)
Edad <sup>2</sup>	-0.0002*** (0.0001)	0.0002*** (0.00003)	-0.001*** (0.00004)
Nivel Educativo Medio	-0.112*** (0.037)	-0.337*** (0.021)	0.267*** (0.027)
Nivel Educativo Alto	0.131** (0.052)	-0.436*** (0.029)	0.406*** (0.033)
Constante	-2.672*** (0.154)	-1.280*** (0.080)	-3.888*** (0.112)
Observaciones(n)	26,688	110,337	110,180
Log Likelihood	-13,238.230	-45,180.400	-31,683.570
Akaike Inf. Crit.	26,504.470	90,388.810	63,395.150

Nota:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

medio tiene una mayor probabilidad de ser migrante en relación al grupo base, misma relación se repite si se considera una persona con nivel educativo alto (educación superior o universitaria), en donde para el cluster 1 y 3 la probabilidad de ser migrante aumenta en comparación con el grupo base, generando una expulsión de personas con elevado nivel educativo, lo cual es inverso en el cluster 2, el cual la probabilidad de migrar se reduce en comparación con el grupo base si se posee un nivel educativo superior o universitario.

## **Factores determinantes de atracción de los migrantes**

El deseo de migrar es distinto a la acción de migración, entre el deseo y el hecho existen múltiples barreras y condicionantes que impiden su concreción, es por ello que el análisis de los factores que determinarán la atracción de las personas hacia los distintos clusters se realiza con respecto a la muestra en el período de relevancia de personas que efectivamente migraron.

La distinción se realiza en torno a los factores que determinan que un migrante elija como destino alguno de los clusters, se busca analizar si existen diferencias significativas en las características de un migrante dependiendo del cluster que elige como destino de migración.

El sexo es un determinante significativo para los migrantes a la hora de elegir su lugar de destino, los migrantes hombres tienen una mayor probabilidad de seleccionar como lugar de destino a las provincias del cluster 1 y 2, lo cual es inverso para las provincias del cluster 3, el cual tiene una mayor probabilidad de selección si la migrante es mujer. La edad de los migrantes tiene un efecto muy diferenciado dependiendo del cluster que se esté considerando como destino, para los migrantes que eligen el cluster 1, la edad es un determinante negativo, el cual genera que cuanto mayor sea la edad del migrante, menor sea la probabilidad que elija el cluster 1 como destino, para el cluster 2 y 3 la relación con la edad de los migrantes es cuadrática, siendo que ante una mayor edad se reduce la probabilidad de elegir como destino de migración el cluster 2, pero cada año que se suma tiene un efecto negativo cada vez menor, efecto totalmente inverso tiene la edad en los que eligen como destino de migración el cluster 3, para los cuales la edad es un factor que afecta positivamente a la probabilidad de migrar hacia este cluster, pero cada año que se suma tiene un efecto positivo cada vez menor.

La incidencia en la pobreza es un factor socioeconómico que es relevante a la hora de que un migrante decida su cluster de destino, la probabilidad de que una persona pobre elija como destino los cluster 1 y 2, *ceteris paribus*, es mayor que la de una persona NO-pobre, sin embargo, esta relación se invierte si se considera como destino el cluster 3, en donde la probabilidad de que un migrante lo elija como destino es menor en el caso de una persona pobre.

La recepción de subsidios o ayudas por parte del gobierno, instituciones, iglesias, etc, tiene un impacto significativo en la probabilidad de elegir como destino alguno de los tres clusters, sin embargo el patrón es distinto dependiendo de cluster que se tome como destino, la recepción de un subsidio disminuye la probabilidad de elegir como destino el cluster 1 y 3, mientras que en el cluster 2 el efecto es contrario, en donde la recepción de esta transferencia económica aumenta la probabilidad de elegir como destino este cluster.

La situación y cualidad ocupacional es un factor determinante para la elección del cluster de destino para los migrantes, para los cluster 1 y 2 la dirección del efecto de este determinante es similar, la probabilidad de elegir como destino de migración alguno de estos clusters se reduce en presencia de un trabajo de baja calificación en relación a una persona que está desocupada o inactiva (en adelante grupo base), y este efecto se invierte si la condición de ocupación de la persona es en un trabajo de alta calificación, en donde la probabilidad de elegir como destino de migración el cluster 1 y 2 aumenta en comparación con el grupo base, lo cual indica que un migrante con un trabajo de calificación alta tiene, *ceteris paribus*, mayor probabilidad en comparación con el grupo base de elegir como destino a provincias que pertenezcan al cluster 1 y 2. Para las provincias del cluster 3 encontramos una situación disímil, en donde la probabilidad de elegir como destino de migración este cluster aumenta en presencia de un trabajo de baja calificación en relación a una persona que está desocupada o inactiva (en adelante grupo base), y este efecto se invierte si la condición de ocupación de la persona es en un trabajo de alta calificación, lo que hace menos probable (*ceteris paribus*, en relación con el grupo base) que un migrante con un trabajo de calificación alta elija como provincia de destino a alguna de las pertenecientes al cluster 3.



La organización familiar vuelve a jugar un papel relevante en la elección de los migrantes en torno al cluster de destino, en este caso, ser soltero/a con hijos/as aumenta la probabilidad de que un migrante sea atraído como lugar de destino por los cluster 1 y 2 en comparación con una persona soltera sin hijos (en adelante grupo base), mientras que para el cluster 3 el efecto es inverso, donde ser soltero con hijos disminuye la probabilidad de elegir como destino este cluster, en relación con el grupo base.

Los migrantes que están en pareja (casado o unido) SIN hijos tienen una mayor probabilidad de elegir como destino de migración el cluster 1 y 3, en comparación el grupo base, situación inversa se configura si se considera como destino alguna de las provincias del cluster 2, en donde estar en pareja SIN hijos disminuye la probabilidad de elegir como destino este cluster en comparación con el grupo base. En el caso de las personas que están casadas CON hijos, tienen mayor probabilidad de elegir como destino los cluster 1 y 2 en comparación con el grupo base, situación inversa se da en los migrantes casados CON hijos que ven reducida su probabilidad de elegir como destino el cluster 3 en comparación con una persona soltera sin hijos.

Un migrante que es propietario de su vivienda tiene menos probabilidad de elegir como provincia de destino a alguna de las que pertenecen al cluster 1 y 3, mientras que este efecto es contrario para las provincias que pertenecen al cluster 2, en donde la probabilidad de ser un migrante que elige como cluster de destino este mismo aumenta en el caso de ser propietario de la vivienda.

Para finalizar el análisis de los factores de atracción de los migrantes hacia los cluster de destino, se tiene en cuenta el efecto de la educación del migrante en la probabilidad de elegir alguno de los tres cluster, un migrante con nivel educativo medio (secundaria completa) tiene mayor probabilidad de haber elegido al cluster 2 como su cluster de destino en relación a un migrante con un nivel educativo bajo (en adelante grupo base), sin embargo, para el cluster 3 esta probabilidad es inversa, en donde un migrante con nivel educativo medio tiene una menor probabilidad de elegir como destino a alguna de las provincias que componen este cluster en relación al grupo base. Un migrante con un nivel educativo alto tiene mayor probabilidad de elegir como destino el cluster 1 u 2 en relación al grupo base, y esta probabilidad se reduce si se considera como destino el cluster 3, en el cual un migrante con un nivel educativo alto tiene menor probabilidad de elegir como destino alguna de las provincias que componen el cluster 3 en relación al grupo base.