



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO  
PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS MATEMÁTICAS Y DE LA  
ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA

APLICACIÓN DE ANÁLISIS DE CORRESPONDENCIAS A EMPRESAS INNOVADORAS EN  
MÉXICO

TESINA PARA OPTAR POR  
EL GRADO DE  
ESPECIALISTA EN ESTADÍSTICA APLICADA

PRESENTA:  
ANTONIO CHIAPA ZENON

ASESOR:  
M. EN C. JOSÉ SALVADOR ZAMORA MUÑOZ, FACULTAD DE CIENCIAS UNAM

MÉXICO, D.F. 2015



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## Índice

<b>ÍNDICE</b>	<b>1</b>
<b>ÍNDICE DE GRÁFICAS Y TABLAS</b>	<b>1</b>
<b>INTRODUCCIÓN</b>	<b>3</b>
<b>ANTECEDENTES Y CONSIDERACIONES METODOLÓGICAS</b>	<b>4</b>
<b>CONCEPTOS IMPORTANTES DEL ANÁLISIS DE CORRESPONDENCIAS (AC)</b>	<b>8</b>
PERFIL	9
ESPACIO DE PERFIL	12
MASA	15
INERCIAS TOTAL	16
<b>EXPLORACIÓN DE DATOS</b>	<b>19</b>
TIPO DE VINCULACIÓN Y ÁREA DE CONOCIMIENTO	20
ÁREA DE CONOCIMIENTO Y GRADO ACADÉMICO DE LOS RECURSOS HUMANOS EXTERNOS	23
TIPO DE VINCULACIÓN Y ESTADOS DE LA REPÚBLICA	25
<b>ANÁLISIS DE CORRESPONDENCIAS</b>	<b>29</b>
ÁREA TECNOLÓGICA Y EL TIPO DE VINCULACIÓN	29
ÁREA DE CONOCIMIENTO Y GRADO ACADÉMICO	33
ESTADO Y TIPO DE VINCULACIÓN	37
<b>CONCLUSIONES</b>	<b>43</b>
<b>BIBLIOGRAFÍA</b>	<b>45</b>

## Índice de gráficas y tablas

GRÁFICA 1 POSICIÓN DE LOS PERFILES DE RENGLÓN Y PERFIL PROMEDIO EN UN ESPACIO DE TRES DIMENSIONES	13
GRÁFICA 2 REPRESENTACIÓN EN DOS DIMENSIONES DEL GRÁFICO 1	13
GRÁFICA 3 FRECUENCIAS DE PROYECTOS POR TIPO DE VINCULACIÓN Y ÁREA TECNOLÓGICA	22
GRÁFICA 4 PERFILES DE ÁREA PECITI	23
GRÁFICA 5 FRECUENCIAS DE PROYECTOS POR TIPO DE VINCULACIÓN Y GRADO ACADÉMICO	24
GRÁFICA 6 PERFILES DEL GRADO ACADÉMICO DE LOS RH	25
GRÁFICA 7 FRECUENCIAS DE PROYECTOS DE TIPO DE VINCULACIÓN Y ESTADOS	27
GRÁFICA 8 PERFILES DE ESTADO	28
GRÁFICA 9 REPRESENTACIÓN GRÁFICA DEL AC	31
GRÁFICA 10 REPRESENTACIÓN GRÁFICA DEL AC	35
GRÁFICA 11 REPRESENTACIÓN GRÁFICA DEL AC	39
TABLA 1 TAXONOMÍA DE LA VINCULACIÓN	6
TABLA 2 EJEMPLO DE FRECUENCIAS	10
TABLA 3 EJEMPLO PERFIL PROMEDIO PARA EL CASO DE RENGLONES	10
TABLA 4 EJEMPLO DE PERFIL PROMEDIO PARA EL CASO DE LAS COLUMNAS	11
TABLA 5 MASAS DE RENGLÓN	15
TABLA 6 MASAS DE COLUMNA	15
TABLA 7 VARIABLES UTILIZADAS	19
TABLA 8 FRECUENCIAS ABSOLUTAS DE ÁREA PECITI Y TIPO DE VINCULACIÓN	21

TABLA 9 FRECUENCIAS ABSOLUTAS DE ÁREA PECITI Y RECURSOS EXTERNOS	24
TABLA 10 FRECUENCIAS ABSOLUTAS DE ESTADO Y TIPO DE VINCULACIÓN	26
TABLA 11 ÁREA PECITI Y TIPO DE VINCULACIÓN MODIFICADA	29
TABLA 12 PORCENTAJE DE INERCIA EXPLICADA	30
TABLA 13 MASA E INERCIA POR RENGLÓN	32
TABLA 14 MASA E INERCIA POR COLUMNA	32
TABLA 15 CONTRIBUCIÓN DE LA INERCIA POR CELDA	33
TABLA 16 ÁREA PECITI Y GRADO ACADÉMICO MODIFICADA	33
TABLA 17 PORCENTAJE DE INERCIA EXPLICADA	34
TABLA 18 MASA E INERCIA POR RENGLÓN	36
TABLA 19 MASA E INERCIA POR COLUMNA	36
TABLA 20 CONTRIBUCIÓN A LA INERCIA POR CELDA	37
TABLA 21 PORCENTAJE DE INERCIA EXPLICADA	38
TABLA 22 MASA E INERCIA POR RENGLÓN	40
TABLA 23 MASA E INERCIA POR COLUMNA	41
TABLA 24 CONTRIBUCIÓN A LA INERCIA POR CELDA	42
TABLA 25 PATRONES DE COMPORTAMIENTO ENTRE EL GRADO DE COMPLEJIDAD DE VINCULACIONES Y GRADO ACADÉMICO	45

## Introducción

Esta tesina forma parte de la primera fase del proyecto Diagnóstico sobre la relación de Oferta-Demanda entre los posgrados nacionales de excelencia (PNPC) y el sector productivo que ha sido apoyado por los fondos de innovación de Secretaría de Economía y el Consejo Nacional de Ciencia y Tecnología.

El objetivo de esta tesina es generar hipótesis sobre la dinámica de la demanda del conocimiento del sector productivo, a través de un estudio retrospectivo en el cual se analiza la información de las fichas de inscripción y técnicas de los proyectos. El universo de estudio es de 571 fichas de empresas que han sido apoyadas por el Fondo de Innovación Tecnológica (FIT) y el programa AVANCE de 2007 a 2013. Se analizaron las relaciones entre (i) tipo de vinculación, (ii) las áreas de conocimiento, (iii) el grado académico de los recursos humanos externos contratados y (iv) el estado en el que se desarrollaron los proyectos.

Para tal motivo, se decidió aplicar la técnica multivariada de análisis de correspondencias (AC) y la estadística Ji-Cuadrada a fin de establecer las relaciones que existen entre los puntos (i, ii, iii y iv) y generar supuestos sobre la dinámica de los proyectos.

Este trabajo consta de un apartado de antecedentes y consideraciones metodológicas, un apartado sobre los conceptos de la técnica de AC, un capítulo de exploración de datos y el análisis de correspondencia. El tercer capítulo muestra el análisis descriptivo de las tablas de contingencia que se utilizaron. Finalmente, se presentan los resultados del uso de la prueba de la Ji-Cuadrada y el análisis de correspondencias.

## Antecedentes y consideraciones metodológicas

El sector productivo requiere recursos humanos altamente capacitados acorde con las necesidades de sus procesos productivos, productos, las tecnologías que utiliza, etcétera. Se sabe que una parte de esos recursos son formados en el extranjero, no obstante existen en el país diversos programas con capacidad para satisfacer su demanda, pero no se sabe en qué medida los posgrados del país conocen las características de la demanda por parte del sector productivo y en qué medida logran satisfacerla.

El PNPC es un programa que fortalece institucionalmente los posgrados de calidad del país; la mayoría son programas académicos, no profesionalizantes y con escasa o nula vinculación con el sector productivo. En este sentido, se considera que se desaprovecha la capacidad instalada de varios posgrados para atender necesidades de la industria, pues los recursos humanos no se forman en relación a las demandas específicas del sector productivo regional o nacional.

Por tal motivo, se desarrolló el programa Diagnóstico sobre la relación de Oferta-Demanda entre los posgrados nacionales de excelencia (PNPC) y el sector productivo. La meta de este diagnóstico es analizar la posible convergencia que existe entre la oferta académica de posgrado y los requerimientos de personal altamente especializado que han solicitado las empresas que recurren a sus fondos, a saber, el Fondo de Innovación Tecnológica (FIT)<sup>1</sup> y el programa AVANCE<sup>2</sup>.

Para lograr dicho objetivo el programa se ha dividido en 4 fases. La primera etapa está abocada a analizar la información recopilada durante la realización de los proyectos. Cada uno de los proyectos llena una ficha técnica de inscripción y fichas de reporte

---

<sup>1</sup> Fideicomiso público creado entre la Secretaría de Economía, la Subsecretaría para la Pequeña y Mediana Empresa y el Consejo Nacional de Ciencia y Tecnología (CONACYT), con el propósito de apoyar a las

<sup>2</sup> Es un programa creado para impulsar la creación de negocios basados en la explotación de desarrollos científicos y/o desarrollos tecnológicos. Cabe aclarar que este es un programa que no se encuentra más en las convocatorias, pero la evaluación resulta importante porque otra de las metas del Consejo es modificarlo sustancialmente y lanzarlo nuevamente al mercado.

técnico. En la primera se describen las actividades generales de la empresa, monto solicitado, nombres de los responsables técnicos y administrativos, etcétera. La segunda es una bitácora para reportar los hallazgos, monitorear el proceso de desarrollo, informar sobre el gasto de los fondos.

Uno de los problemas sobre la calidad de la información obtenida en esos documentos es su enfoque administrativo. Los aspectos sobre gestión del conocimiento i.e. impacto en las capacidades de desarrollo tecnológico de los proyectos, requerimientos cognitivos no satisfechos, no son tomados en cuenta. De tal manera, la información relativa a la dinámica del conocimiento especializado es muy poca.

Las variables críticas que se obtienen de las fichas son: (i) la vinculación con institutos de educación superior, universidades y otras empresas para el desarrollo del proyecto; (ii) el área de conocimiento a la cual pertenece el proyecto; (iii) qué tipo de recursos externos humanos especializados utilizaron y (iv) la entidad federativa.

Con respecto a la variable vinculación se encontró que en las fichas se reportan actividades sumamente diversas como lo son:

1. Prácticas profesionales.
2. Pruebas laboratorio.
3. Control de calidad.
4. Capacitación especializada.
5. Asistencia técnica.
6. Prácticas profesionales.
7. Tesis s/proyectos.
8. Desarrollo de componentes para proyectos de I+D.
9. Realización de partes de proyectos.
10. Transferencia de tecnología.
11. Prototipos.
12. Co-desarrollo de tecnología.

13. Proyectos Co-innovación en procesos.
14. Proyectos Co-innovación en productos.
15. Co-patentes.

Para condensar dichas actividades en categorías útiles para una tabla de contingencia se construyó una taxonomía que diferencia su grado de complejidad. A continuación se describen las tres categorías usadas y las actividades incorporadas en cada una de ellas.

**Tabla 1 Taxonomía de la vinculación**

<b>Puntual</b>	<b>Modular</b>	<b>Integral</b>
Prácticas profesionales	Prácticas profesionales	Prototipos
Pruebas laboratorio	Tesis s/proyectos	Co-desarrollo de tecnología
Control de calidad	Desarrollo de componentes para proyectos de I+D	Proyectos Co-innovación en procesos
Capacitación especializada	Realización de partes de proyectos	Proyectos Co-innovación en productos
Asistencia técnica	Transferencia de tecnología	Co-patentes

Las áreas de conocimiento se establecieron según la categorización del programa especial de ciencia y tecnología (PECITI) del 2008-2012. Éstas son:

- (i) Biotecnología.
- (ii) Medicina.
- (iii) Energía.
- (iv) Medio Ambiente.
- (v) Tecnologías industriales de fabricación.
- (vi) Materiales.
- (vii) Nanotecnología.
- (viii) Tecnologías de la información y las telecomunicaciones.
- (ix) Matemáticas aplicadas y modelación.

Sobre los recursos humanos especializados que contrataron las empresas se obtuvo información sobre el grado académico (licenciatura, maestría o doctorado) y el área de conocimiento.



El objetivo de esta tesina es coadyuvar al análisis de esta información para generar hipótesis sobre la demanda y dinámica del conocimiento especializado del sector productivo.

El universo de estudio es de 571 fichas de empresas que han sido apoyadas por el Fondo de Innovación Tecnológica (FIT) y el programa AVANCE de 2007 a 2013. Se analizaron las relaciones entre las siguientes variables: (i) tipo de vinculación, (ii) las áreas de conocimiento, (iii) el grado académico de los recursos humanos externos contratados y (iv) el estado en el que se desarrollaron los proyectos.

En este sentido, este trabajo es de corte retrospectivo y la información obtenida es un censo de las empresas apoyadas en el periodo mencionado. No obstante, la técnica de AC no tiene validez interna ni externa en tanto técnica descriptiva. Su objetivo no es la prueba de hipótesis sino la exploración de los datos. De acuerdo a los criterios de clasificación de las investigaciones se considera un estudio descriptivo, transversal, retrospectivo y observacional.

## Conceptos importantes del Análisis de Correspondencias (AC)

El análisis de correspondencias es una herramienta útil para explorar las relaciones entre los distintos niveles de una tabla de contingencia y generar hipótesis que puedan ser útiles en las siguientes etapas del Diagnóstico.

Esta técnica mide la asociación o independencia de las dimensiones de las variables que se encuentran en una tabla de contingencia. A diferencia de la Ji-cuadrada, que establece el grado de asociación *global* entre las variables, el AC es una excelente herramienta analítica para evaluar la asociación entre los renglones y las columnas que la conforman.

Cabe mencionar que el AC se divide en simple y múltiple. El primero analiza la asociación entre dos variables mientras que el segundo puede incorporar más de dos. Para este trabajo se utiliza únicamente el primero.

La notación de una tabla de contingencia de dos por dos se presenta a continuación:

	<b>Y<sub>1</sub></b>	<b>Y<sub>2</sub></b>	<b>Total</b>
<b>X<sub>1</sub></b>	$n_{11}$	$n_{12}$	$n_{1+}$
<b>X<sub>2</sub></b>	$n_{21}$	$n_{22}$	$n_{2+}$
<b>Total</b>	$n_{+1}$	$n_{+2}$	$n_{++}$

Dónde:

$x_1, x_2$  = los niveles de la variable  $X$

$y_1, y_2$  = los niveles de la variable  $Y$

$n_{11}, n_{12}, n_{21}, n_{22}$ , = el número de elementos de la tabla de contingencia que se encuentran en el renglón  $i$ -ésimo y en la columna  $j$ -ésima. Donde la  $i$  corresponde al primer nivel de cada variable y la  $j$  al segundo.

$n_{1+}, n_{2+}$  = el número total de elementos que se encuentran en el primer y segundo renglón (o nivel de la variable  $x$ ).

$n_{+1}, n_{+2}$  = el número total de elementos que se encuentran en la primer y

segunda columna (o nivel de la variable y).

$n_{++}$  = el número de elementos totales de la muestra.

Otra notación importante que se usará en adelante:

$p_{ij} = n_{ij}/n_{++}$  = la probabilidad de encontrarse en el renglón *i-ésimo* y la columna *j-ésima*.

$r_i = n_{i+}/n_{++} = p_{i+}$  = la masa del renglón *i-ésimo*, el vector se denota **r**

$c_j = n_{+j}/n_{++} = p_{+j}$  = la masa de la columna *j-ésima*, el vector se denota **c**

$a_{ij} = n_{ij}/n_{i+}$  = el *j-ésimo* del perfil del renglón.

$b_{ij} = n_{ij}/n_{+j}$  = el *i-ésimo* elemento del perfil de columna.

Mientras la mayoría de las herramientas estadísticas se concentran en problemas en los que los datos son ajustados y comparados con un modelo teórico o un conjunto de hipótesis preconcebidas, el AC permite generar hipótesis que pueden ser evaluadas en subsecuentes etapas de investigación (Greenacre, 2007). En este sentido, el AC es una técnica descriptiva con la cual no se pueden hacer inferencias y carece de validez externa.

A continuación se exponen los conceptos que serán utilizados en la interpretación de los resultados, a saber, (i) perfil y espacio de perfil, (ii) masa, (iii) promedio ponderado, (iv) distancia ji-cuadrada e (v) inercia.

## Perfil

El concepto de perfil es fundamental en el AC. Este se define como aquel conjunto de frecuencias relativas de las dimensiones de las variables en una tabla de contingencia.

Por ejemplo, en la siguiente tabla se aprecian las frecuencias de 86 días de viajes al extranjero de una persona. Éstos se encuentran divididos en vacaciones, medios días y días completos<sup>3</sup>. La tabulación cruzada con respecto a los países se muestra a continuación.

**Tabla 2 Ejemplo de frecuencias**

País	Vacaciones	Medio día	Día completo	Total columna
Noruega	6	1	11	18
Canadá	1	3	11	15
Grecia	4	25	0	29
Francia / Alemania	2	2	20	24
<b>Total renglón</b>	<b>13</b>	<b>31</b>	<b>42</b>	<b>86</b>

Fuente: Tomado del libro Análisis de correspondencias en la práctica. Greenacre, 2007.

Estos conjuntos o vectores de frecuencias relativas tienen características geométricas especiales porque la suma de los elementos en dichos conjuntos es 1 o 100%. Por ejemplo, en la tabla 2 el perfil del renglón 1 estaría dado por el siguiente conjunto de frecuencias relativas y cuya suma es uno.

$$\text{NORUEGA } [6/18, 1/18, 11/18] = 1$$

De acuerdo a la notación presentada anteriormente cada uno de los puntos de ese vector es el resultado de:

$$a_{ij} = n_{ij}/n_{i+}$$

En la siguiente tabla se muestran estos perfiles para cada nivel de la variable país:

**Tabla 3 Ejemplo perfil promedio para el caso de renglones**

País	Vacaciones	Medio día	Día completo
Noruega	.33	.06	.61
Canadá	.07	.2	.73
Grecia	.14	.86	0

<sup>3</sup> Los medios días y días completos son aquellos viajes de trabajo en los cuáles solo se estuvo doce y 24 horas respectivamente en el país de visita,

<b>Francia / Alemania</b>	.08	.08	.83
<b>Perfil promedio de renglón</b>	.15	.36	.49

Fuente: Tomado del libro Análisis de correspondencias en la práctica. Greenacre, 2007.

Como se puede observar existe un último renglón denominado perfil promedio de renglón. Este se obtiene de la misma manera pero con base en la suma total de los renglones y cuya suma es uno.

$$\text{Perfil promedio de renglón [13/86, 31/86, 42/86]} = 1$$

Como se puede apreciar en la tabla 2 los 6 días en Noruega para la columna de vacaciones, comparados con los 4 días en Grecia, son evaluados en relación al número de días para cada país. En este sentido, las frecuencias relativas resultan bastante diferentes. Para el primer caso es igual a .33 (6 de 18), mientras que para Grecia es sólo de .14 (4 de 29).

De manera análoga, el perfil de columna para el caso de vacaciones, cuya suma es uno, estaría dado por:

$$\text{VACACIONES [6/13, 1/13, 4/13, 2/13]} = 1$$

Y su perfil promedio por:

$$\text{Perfil promedio de columna [18/86, 15/86, 29/86, 24/86]} = 1$$

En la siguiente tabla se presentan los perfiles por columna.

**Tabla 4 Ejemplo de perfil promedio para el caso de las columnas**

<b>País</b>	<b>Vacaciones</b>	<b>Medio día</b>	<b>Día completo</b>	<b>Perfil promedio</b>
<b>Noruega</b>	.46	.03	.26	.21
<b>Canadá</b>	.08	.10	.26	.17
<b>Grecia</b>	.31	.81	0	.34
<b>Francia / Alemania</b>	.15	.07	.48	.28

Fuente: Tomado del libro Análisis de correspondencias en la práctica. Greenacre, 2007.

Al dividir los perfiles entre los perfiles promedio tendríamos que, para el caso de los renglones, la tasa de días de vacaciones pasados en noruega es 2.2 (.33/.15), a esto se le denomina tasa de contingencia. Si se obtuviera la tasa de contingencia para el caso de los perfiles columna se obtendría el mismo resultado. Por ejemplo, para el caso de Noruega se obtuvo un perfil de .46 que dividido en su perfil promedio .21 nos da un resultado de 2.2. Es decir, sin importar que tipo de perfil se usa (sea por renglones o por columnas) llegamos al mismo resultado. A la técnica de AC que trata a los renglones y las columnas en el mismo modo (tasas de contingencia iguales) se le denomina simétrica, mientras que a la que otorga un peso diferente a cualquiera de las dos se le llama asimétrica. Como es el caso de este estudio sólo nos enfocaremos a la primera.

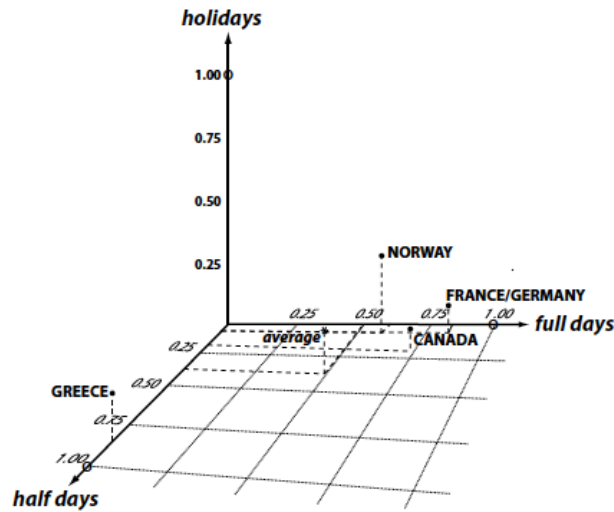
### Espacio de perfil

Este se define como el plano de  $m^4$  dimensiones en el cual se encuentran los perfiles, que a su vez se conforman por  $m$  elementos y cuya suma es igual a 1. Los perfiles ocupan un sub espacio restringido a  $m-1$  dimensiones que se denomina simplex. Por ejemplo, si se grafican los perfiles de la tabla 3 se obtiene la gráfica 1. Las columnas se encuentran en cada uno de los ejes (x,y,z) y en el espacio tridimensional se grafican los 4 países así como el perfil promedio. De tal manera, para Noruega se tiene el punto con las coordenadas (.33,.06,.61); el perfil promedio se encuentra en las coordenadas (.15, .36, .49). Así, cada uno de los puntos se encuentran en el espacio de un triángulo que une los puntos extremos [0,0,1],[0,1,0],[1,0,0]. A este triángulo equilátero y a sus tres puntos extremos se le denomina puntos vértice.

---

<sup>4</sup>  $m$  es el valor que adquiere un plano de cualquier dimensión. Por ejemplo, para una tabla de dos por dos se asume que su representación es bidimensional y por tanto  $m$  es igual a 2.

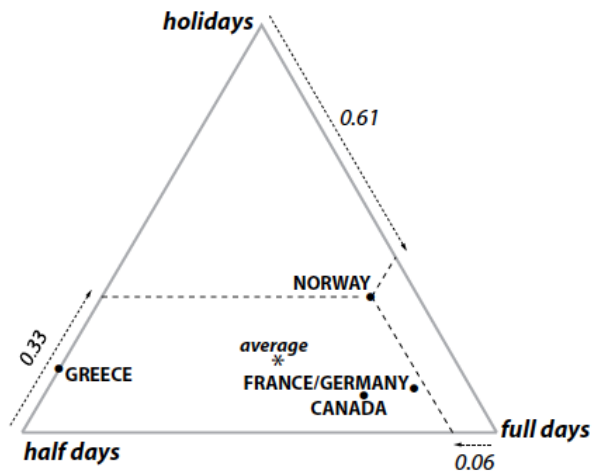
## Gráfica 1 Posición de los perfiles de renglón y perfil promedio en un espacio de tres dimensiones



Fuente: Tomado del libro Análisis de correspondencias en la práctica. Greenacre, 2007.

Una representación bidimensional del espacio de perfiles quedaría de la siguiente forma:

## Gráfica 2 Representación en dos dimensiones del gráfico 1



Fuente: Tomado del libro Análisis de correspondencias en la práctica. Greenacre, 2007.

Este sistema coordenado triangular sólo puede ser usado para perfiles con 3 elementos, pero la idea puede ser generalizada fácilmente a  $n$  perfiles en cuyo caso se

denomina sistema coordinado baricéntrico. Así, en la gráfica 2 la lectura de la posición relativa de los perfiles es más clara. Se observa una mayor cercanía de los perfiles de noruega, Francia/Alemania y Canadá con el perfil de días completos mientras que Grecia está más próxima a la categoría de medios días. Es evidente que el perfil promedio se encuentra muy cercano a dichos perfiles y muy alejado del perfil de vacaciones.

Otro concepto esencial es el de centroide o promedio ponderado. Los puntos en la gráfica 2 (Noruega, Canadá, Grecia y Francia/Alemania) pueden verse como promedios ponderados. Por ejemplo, para el caso de Noruega el promedio ponderado se obtendría sumando las frecuencias absolutas para cada una de las categorías de las columnas multiplicados por cada uno de los elementos del perfil. Por ejemplo, para el caso de Noruega se tendría:

Centroide= (frecuencia días de vacaciones\* perfil de renglón 1) + (frecuencia medios días \* perfil de renglón 2) + (frecuencia días completos \* perfil de renglón 3)

$$6 * .33 + 1 * .06 + 11 * .61$$

De esta manera se esperarí que en el sistema coordinado triangular la posición del promedio ponderado para Noruega estuviera más cercano a los días completos. De la misma manera se podría obtener un promedio ponderado para el perfil promedio por renglón de este ejemplo:

Centroide= (frecuencia total renglón vacaciones \* perfil promedio de renglón) + (frecuencia total renglón medios días \* perfil promedio de renglón) + (frecuencia total renglón días completos \* perfil promedio de renglón)

$$13 * .15 + 31 * .36 + 42 * .49$$



## Masa

A las frecuencias marginales por renglón también se le conoce como masas del renglón. Por tanto, para este caso, las masas se obtienen dividiendo los días de viaje en cada país entre el total de días.

$$r_i = n_{i+}/n$$

A continuación se presenta el cálculo para todos los países de nuestro ejemplo:

**Tabla 5 Masas de renglón**

País	Vacaciones	Medio día	Día completo	Masas de renglón
Noruega	.33	.06	.61	0.21
Canadá	.07	.2	.73	0.17
Grecia	.14	.86	0	0.34
Francia / Alemania	.08	.08	.83	0.28
<b>Perfil promedio</b>	<b>.15</b>	<b>.36</b>	<b>.49</b>	

Fuente: Tomado del libro Análisis de correspondencias en la práctica.

Para el caso de los perfiles por columna se tiene:

$$c_i = n_{+j}/n$$

**Tabla 6 Masas de columna**

	Vacaciones	Medio día	Día completo	Perfil promedio de columnas
Noruega	0.46	0.03	0.26	0.21
Canadá	0.08	0.10	0.26	0.17
Grecia	0.31	0.81	0.00	0.34
Francia / Alemania	0.15	0.06	0.48	0.28
<b>Masa de columnas</b>	<b>0.15</b>	<b>0.36</b>	<b>0.49</b>	

Fuente: Tomado del libro Análisis de correspondencias en la práctica.

Como se puede observar las masas, de los renglones y las columnas, juegan dos roles: como pesos ponderados y como promedios. En la tabla 4 el perfil promedio de columna es el conjunto de masas del renglón de la tabla 5 y viceversa. De hecho, las

masas son el centroide o promedio ponderado con el que se obtiene el valor del perfil promedio. Por ejemplo, para el caso de los renglones, si se multiplica el perfil del renglón por la masa del renglón:

$$[\text{Noruega}(.33*.21)+ \text{Canadá}(.07*.17)+ \text{Grecia}(.14*.34)+ \text{Francia/Alemania}(.08*.28)]= .15$$

Obtenemos la masa de la columna que es igual a .15

### Inercia total

Otro concepto importante es el de inercia total (IT). La IT se define como la medida de la varianza que hay en los datos de una tabla de contingencia. Geométricamente, la inercia mide qué tan lejos se encuentran los perfiles, de renglón o de columna, de su perfil promedio. El perfil promedio pudiera ser considerado como la hipótesis de homogeneidad (o equidad) de los perfiles.

Por ejemplo, cuando la inercia es baja los perfiles del renglón se encuentran muy cerca de su perfil promedio. Por tanto, existe poca asociación o correlación entre los renglones y las columnas. Por otro lado, mientras la inercia es alta es mayor la asociación que existe entre los perfiles de los renglones y las columnas. Si todos los perfiles fueran idénticos y se encontraran en el mismo punto (su promedio) todas las distancias ji-cuadrada serían cero y por tanto la inercia total también.

La inercia total se obtiene de dividir el resultado de la estadística Ji-cuadrada de una tabla de contingencia entre el número de casos totales de una tabla. Por ejemplo, para la tabla 2 la prueba Ji-cuadrada arroja un valor de 58.9, dividido entre 86, la inercia total sería de: .684

La fórmula para obtener la Ji-cuadrada es:

$$\chi^2 = \sum \frac{(\text{observado}-\text{esperado})^2}{\text{esperado}} \dots \text{Ecuación 1}$$

El resultado de este estadístico se contrasta contra el valor de tablas de una Ji-cuadrada con grados de libertad igual a la multiplicación del número de renglones menos uno por el número de columnas menos uno. Para nuestro ejemplo, los grados de libertad serían 6. Entre mayor es el valor de la  $\chi^2$  más grande es la diferencia entre el valor observado y el valor esperado.

El valor de este estadístico radica en su capacidad de medir una discrepancia en la homogeneidad de los datos. En otras palabras, de medir la heterogeneidad de los perfiles.

Si se re-expresara la fórmula de la ji-cuadrada en términos del perfil esperado y el perfil observado se tendría:

$$Total\ renglón \times \frac{(perfil\ renglón\ observado - perfil\ renglón\ esperado)^2}{perfil\ renglón\ esperado} \dots ecuación\ 2$$

Nota: Se obtiene al dividir cada uno de los términos de la fórmula 1 entre el total del renglón.

Como se mencionó, la IT es igual al valor de la Ji-cuadrada entre el total de las observaciones de una tabla de contingencia, por tanto se tiene:

$$\frac{\chi^2}{86} = \phi^2 = .684$$

$$.21 \times \left[ \frac{(.33-.15)^2}{.15} + \frac{(.06-.36)^2}{.36} + \frac{(.61-.48)^2}{.48} \right] + .17 \times \left[ \frac{(.07-.15)^2}{.15} + \frac{(.2-.36)^2}{.36} + \frac{(.73-.48)^2}{.48} \right] +$$

$$.34 \times \left[ \frac{(.14-.15)^2}{.15} + \frac{(.86-.36)^2}{.36} + \frac{(0-.48)^2}{.48} \right] + .28 \times \left[ \frac{(.08-.15)^2}{.15} + \frac{(.08-.36)^2}{.36} + \frac{(.83-.48)^2}{.48} \right] = .684$$

*ecuación 3*

Lo que da como resultado que cada término de los renglones sea equivalente a la masa del renglón que multiplica a una distancia Ji cuadrada (que está en corchetes). Es por eso que la inercia también se define como:

$$inercia = \sum (masa\ renglón\ iésimo) \times (distancia\ \chi^2\ del\ perfil\ iésimo\ al\ centroide)^2$$

... ecuación 4

Por tanto, la IT es el promedio ponderado (por la masa) del cuadrado de las distancias ji-cuadrada entre los perfiles de columnas o renglones y su perfil promedio. Como todas las técnicas multivariadas, la representación gráfica de la AC implica una reducción de dimensiones. Para hacer esto se usa la inercia como una medida de la varianza o la dispersión geométrica de los puntos<sup>5</sup>.

La línea que mejor ajusta se llama eje principal. Así para el ejemplo que se ha manejado, la IT se encuentra explicada en 2 ejes principales. Si tenemos que la inercia total es .684 y el porcentaje de inercia del primer eje es de 97 % entonces la inercia del primer eje es .663 (.684/.97). La diferencia con respecto a .684 es la inercia del segundo eje principal. A esto se le conoce también como *eigen valor*.

De manera gráfica las relaciones entre los renglones y las columnas de una tabla de contingencia se basan en la idea de representar todas las categorías de las variables e interpretar sus posiciones relativas de los puntos en términos de sus pesos correspondientes a los renglones y las columnas. Aunque hay que tener cuidado con la interpretación, muchas veces son las categorías que más contribuyen a un eje las que le dan nombre al componente. De la misma manera en la que en el análisis factorial la cantidad de varianza explicada por un factor ortogonal es fundamental para el análisis del modelo, en esta técnica la inercia total y la inercia de cada dimensión, son los elementos que ayudan a decidir el ajuste y la interpretación. Es decir, aquellas con mayor valor explican el mayor porcentaje de la varianza, entre mayor porcentaje tengan las dos primeras dimensiones se tendrá una mejor representación en el espacio bidimensional.

---

<sup>5</sup> Para el caso del AC el criterio que se usa para establecer la línea o el plano más cercano a todos los puntos del perfil y los vértices del espacio de perfiles es la suma ponderada de distancias al cuadrado (ponderada por las masas). Para hacerlo, se usa la descomposición por valor singular (DVS). Esta descomposición es a las matrices rectangulares lo que los eigenvalores y eigenvectores son para las matrices cuadradas.

## Exploración de datos

El objetivo de esta sección consiste en analizar las asociaciones estadísticas que existen entre las variables que se definieron como relevantes en el apartado de antecedentes. Éstas son:

- (i) Tipo de vinculación.
- (ii) Temas conocimiento<sup>6</sup> al que pertenecen los proyectos.
- (iii) El grado académico de los recursos humanos externos contratados.
- (iv) El estado en el que se desarrollaron los proyectos.

El siguiente cuadro muestra el nombre, la codificación y el tipo de variables que se utilizan:

**Tabla 7 Variables utilizadas**

Variable	Valores de la variable	Codificación	Tipo de variable
<b>Tipo de vinculación<sup>7</sup></b>	Sin vinculación, Puntual, Modular e Integral.	TipoV	Ordinal
<b>Temas de conocimiento</b>	Sin área Biotecnología (Bio.) Medicina Energía Medio Ambiente (MA) Tecnologías industriales de fabricación (TIF) Materiales Nanotecnología (Nano) Tecnologías de la información y las telecomunicaciones (TICS) Matemáticas aplicadas y modelación (MAM)	Area	Nominal
<b>Grado académico</b>	Sin recursos humanos externos (Sin RE) Sin personal especializado <sup>8</sup>	Grado	Ordinal

<sup>6</sup> Las áreas de conocimiento se establecieron según la categorización del programa especial de ciencia y tecnología (PECITI) del 2008-2012.

<sup>7</sup> Se denomina vinculación a toda relación de trabajo establecida con algún centro de investigación, universidad o empresa para llevar a cabo el proyecto.

	(Sin PE) Al menos un maestro Al menos un doctor		
<b>Estado</b>	Estados de la República	Edo	Nominal

\* Entre paréntesis se escriben las siglas que se utilizarán en las gráficas en las que el nombre completo dificulte la lectura de las gráficas.

Se hizo una tabulación cruzada (por pares) para cada una de las posibles relaciones entre las cuatro variables.

- (i) Tipo de vinculación versus área de conocimiento.
- (ii) Tipo de vinculación versus estado.
- (iii) Tipo de vinculación versus grado académico.
- (iv) Área de conocimiento versus estado.
- (v) Área de conocimiento versus grado académico.
- (vi) Estado versus grado académico.

A continuación se realizó una prueba Ji cuadrada para corroborar la independencia *global* y verificar la pertinencia de ejecutar un análisis de correspondencias. Se presentan sólo aquellas tablas que resultaron significativas en el rechazo de la hipótesis de nulidad de independencia de la prueba.

#### Tipo de vinculación y área de conocimiento

En la tabla 8 se muestran las frecuencias absolutas de los proyectos de acuerdo al tipo de vinculación y el área de conocimiento. La mayor participación la tiene la categoría de *Tecnologías industriales de fabricación* (TIF) con 179 proyectos. Ésta representa casi un tercio del total. La segunda con mayor participación son las *Tecnologías de la información y telecomunicaciones* (TICS) con 112 proyectos (20 %).

---

<sup>8</sup> Aquellos proyectos que no contrataron ni maestros ni doctores. Sólo licenciados, gente con carreras técnicas, diplomados, especialidades, etcétera.

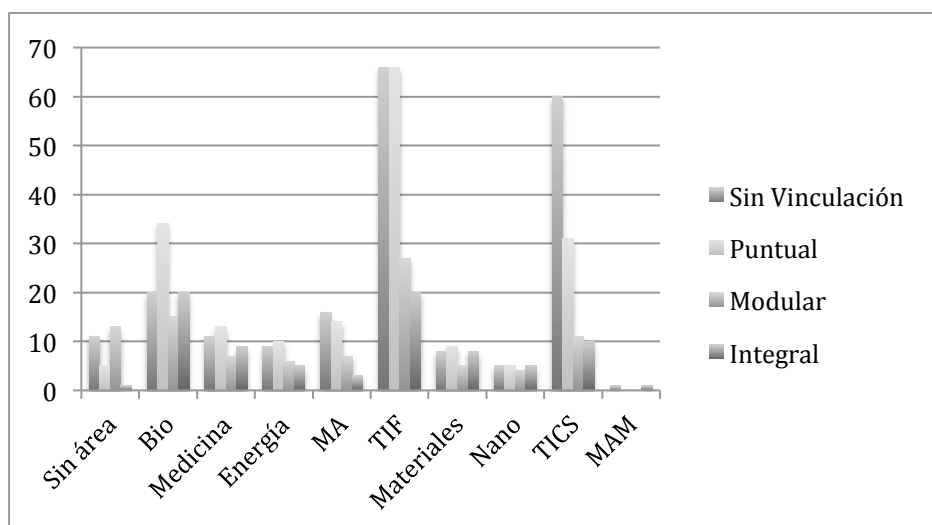
**Tabla 8 Frecuencias absolutas de área Peciti y tipo de vinculación**

	Sin				Total general	%
	Vinculación	Puntual	Modular	Integral		
Sin área	11	5	13	1	30	5.25%
Biotecnología	20	34	15	20	89	15.59%
Medicina	11	13	7	9	40	7.01%
Energía	9	10	6	5	30	5.25%
MA	16	14	7	3	40	7.01%
TIF	66	66	27	20	179	31.35%
Materiales	8	9	5	8	30	5.25%
Nanotecnología	5	5	4	5	19	3.33%
TICS	60	31	11	10	112	19.61%
MAM	1	0	0	1	2	0.35%
Total general	207	187	95	82	571	

De manera gráfica se aprecia que el número de proyectos por tipo de vinculación para cada área de conocimiento es muy diverso, excepto para la categoría de *Materiales* y *Nanotecnología*. Para el área de TICS los proyectos *sin vinculación* y con vinculación *puntual* son 91, pero para las demás categorías la frecuencia es menor a 30. Es decir, el área de tecnologías de la información tiene un sesgo importante hacia actividades que no requieren vinculación o de complejidad tecnológica poco demandante. En este caso se encuentran las áreas de TIF y MA.

Se observa que en ninguna de las áreas la barra *integral* es la de mayor participación. Sólo para el área de *biotecnología* y *TIF* se tiene una frecuencia alta para dicha categoría.

**Gráfica 3 Frecuencias de proyectos por tipo de vinculación y área tecnológica**

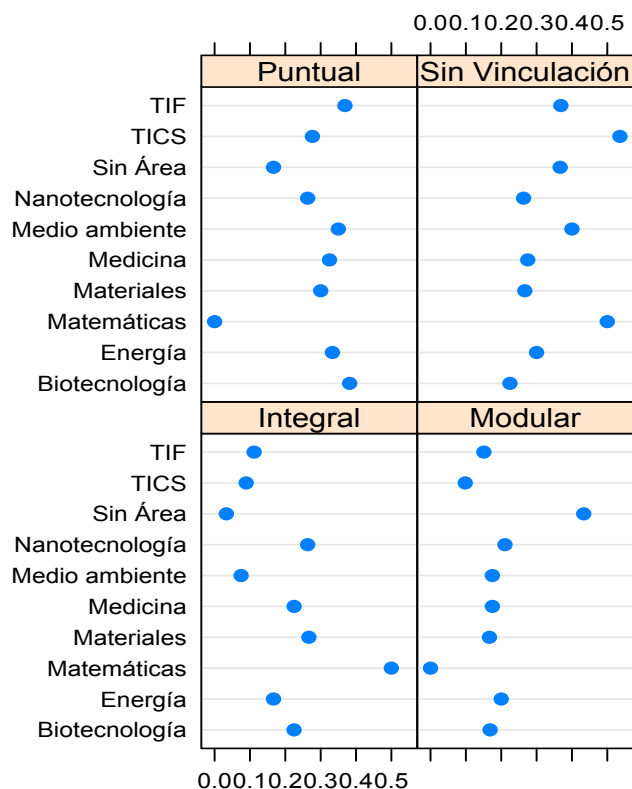


En la siguiente gráfica se muestran los perfiles por renglón. Como se mencionó en el apartado de marco teórico, éstos son la participación relativa de las áreas con respecto al tipo de vinculación. Por ejemplo, las áreas de TIF y de *biotecnología* tienen la frecuencia absoluta más alta de proyectos con vinculación integral (20), no obstante la frecuencia relativa es de 11 % para TIF y de 22 % para *biotecnología*.

En la gráfica 4 se observa una clara tendencia negativa en la participación relativa de los proyectos con vinculación integral para las áreas de TIF, TICS y Medio ambiente. Las áreas de biotecnología, nanotecnología, materiales y energía muestran un comportamiento similar en las categorías de *modular*, *puntual* e *integral*.



Gráfica 4 Perfiles promedio de área PECITI



### Área de conocimiento y grado académico de los recursos humanos externos

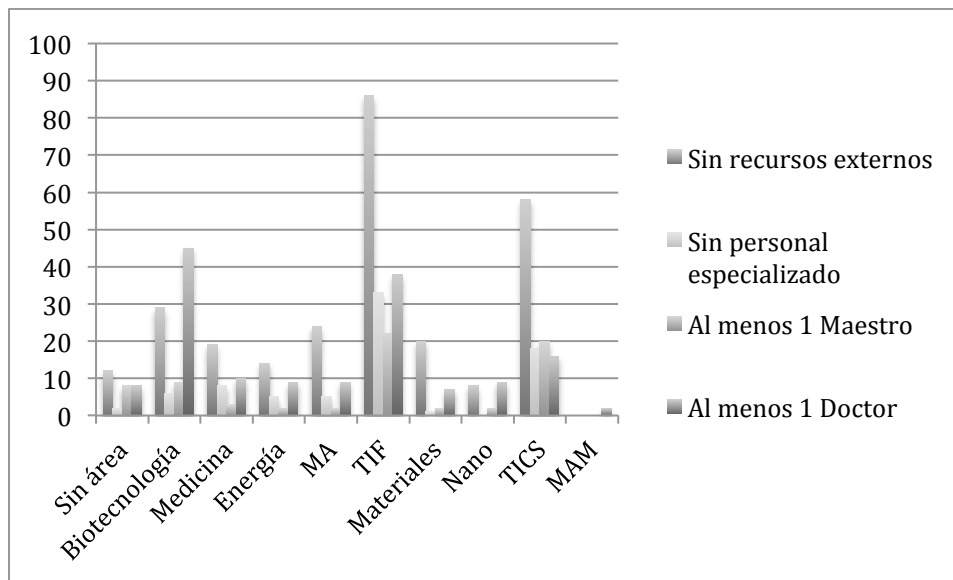
En la siguiente tabla se muestran las frecuencias absolutas del grado académico de los recursos humanos externos (RHE) contratados para el proyecto de acuerdo al área de conocimiento. El área de conocimiento que más contrató doctores fue *Biotecnología* con 45. El segundo renglón con mayor frecuencia es TIF (38) y en tercer lugar TICS (16). Para la categoría de Al menos 1 maestro las áreas que tuvieron frecuencias más altas son TIF y TICS con 22 y 20 respectivamente. Es interesante observar que el área de *Tecnologías industriales de fabricación* también tienen la frecuencia más alta en la categoría de *sin personal especializado* (contrataron licenciados, personal con diplomado y/o especialidad) y *sin recursos externos*.

**Tabla 9 Frecuencias absolutas de área Peciti y recursos externos**

	Sin recursos externos	Sin personal especializado	Al menos 1 Maestro	Al menos 1 Doctor	Total general	%
Sin área	12	2	8	8	30	5.25%
Biología	29	6	9	45	89	15.59%
Medicina	19	8	3	10	40	7.01%
Energía	14	5	2	9	30	5.25%
Medio ambiente	24	5	2	9	40	7.01%
TIF	86	33	22	38	179	31.35%
Materiales	20	1	2	7	30	5.25%
Nanotecnología	8	0	2	9	19	3.33%
TICS	58	18	20	16	112	19.61%
MAM	0	0	0	2	2	0.35%
<b>Total general</b>	<b>270</b>	<b>78</b>	<b>70</b>	<b>153</b>	<b>571</b>	

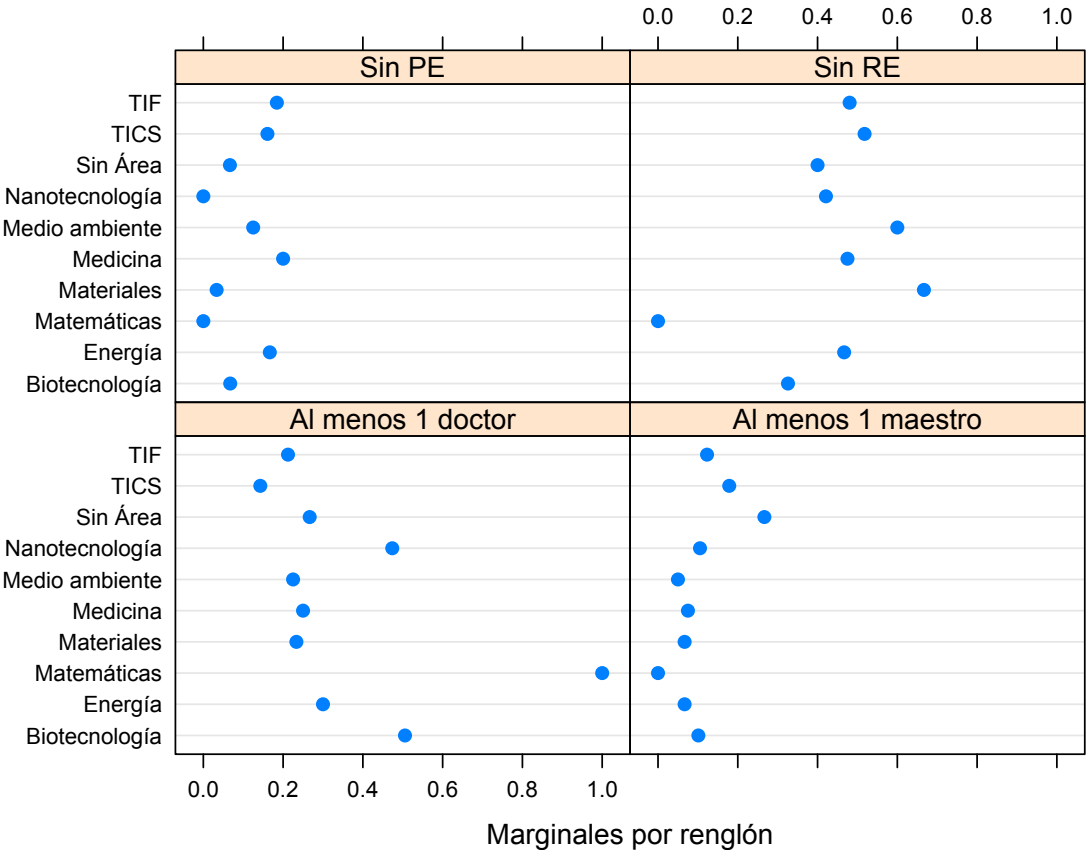
En la gráfica 5 se muestran las frecuencias de la tabla anterior. Es interesante observar que para cada una de las áreas de conocimiento la categoría de sin recursos humanos externos y sin personal especializado son las de mayor frecuencia excepto para biotecnología.

**Gráfica 5 Frecuencias de proyectos por tipo de vinculación y grado académico**



En la gráfica 6 se muestran los perfiles por renglón de la tabla 9. El perfil promedio denota las expectativas *a priori* con respecto al comportamiento del grado académico. Las áreas de conocimiento que se asumen de mayor complejidad tecnológica e.g. nanotecnología, biotecnología y energía con las que mayor número de doctores contratan. No obstante, otras áreas de complejidad similar, como la de materiales y medio ambiente, tienen un comportamiento similar al áreas de conocimiento de menor nivel de complejidad.

**Gráfica 6 Perfiles del grado académico de los RH**



**Tipo de vinculación y estados de la República**

En la siguiente tabla se muestra la frecuencia por estado y tipo de vinculación. Los 571 proyectos se encuentran representados en 26 estados. Como se puede apreciar la lectura de una tabla de contingencia con tantos renglones es muy compleja. De manera general se observa una distribución equitativa y con muy pocos proyectos

para todos los estados excepto el Distrito Federal, Guanajuato, Estado de México, Jalisco y Nuevo León.

**Tabla 10 Frecuencias absolutas de estado y tipo de vinculación**

	Sin vinculación	Puntual	Modular	Integral	Total general	%
Ags	1	1	1	1	4	0.70%
BC	1	4	2	1	8	1.40%
Chih	5	1	3	2	11	1.93%
Coah	10	17	8	7	42	7.36%
DF	66	44	16	20	146	25.57%
Edomex	19	12	6	4	41	7.18%
Gto	16	25	11	10	62	10.86%
Gro	1	0	0	0	1	0.18%
Hgo	0	1	0	0	1	0.18%
Jal	30	17	12	3	62	10.86%
Mich	2	0	4	0	6	1.05%
Mor	6	1	2	0	9	1.58%
NL	26	28	9	15	78	13.66%
Oax	0	0	1	0	1	0.18%
Pue	4	3	4	4	15	2.63%
Qro	3	5	1	1	10	1.75%
QR	2	2	0	0	4	0.70%
SLP	3	2	0	1	6	1.05%
Sin	2	1	2	2	7	1.23%
Son	2	4	1	2	9	1.58%
Tab	2	3	0	0	5	0.88%
Tamps	0	2	2	4	8	1.40%
Tlax	2	2	1	0	5	0.88%
Ver	1	5	2	0	8	1.40%
Yuc	3	7	6	4	20	3.50%
Zac	0	0	1	1	2	0.35%
Total general	207	187	95	82	571	

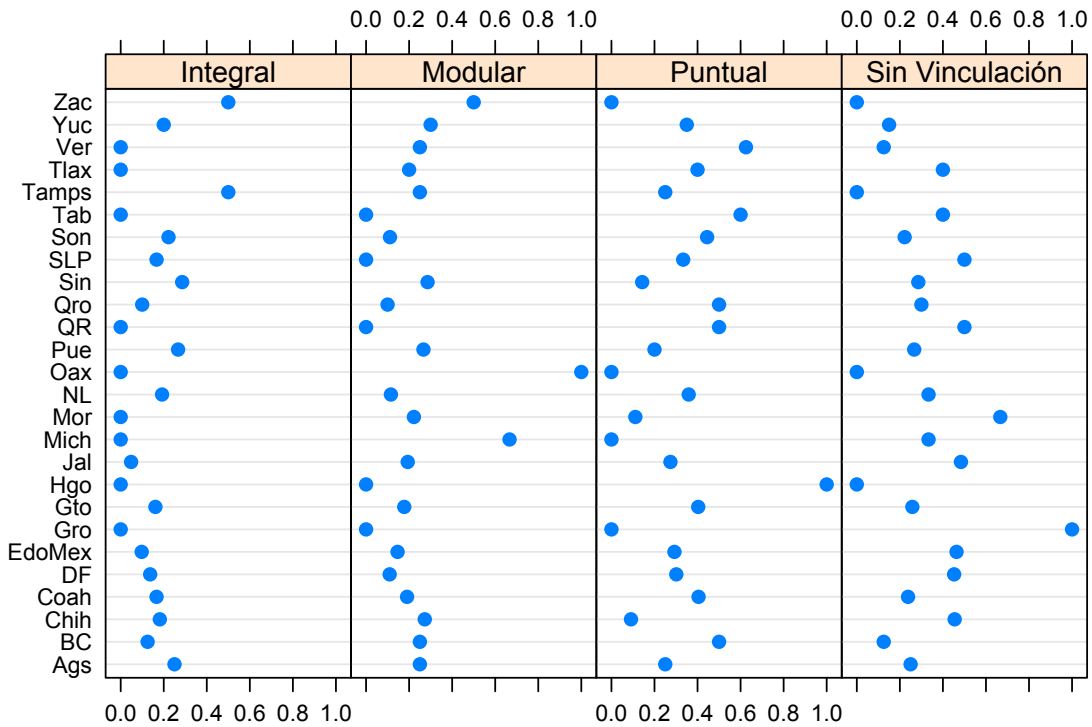
En la siguiente gráfica se presentan las frecuencias de la tabla anterior. A pesar de que el DF tiene la participación más alta de proyectos el número de vinculaciones integrales es equiparable al de estados como Guanajuato y Nuevo León. En general se observa una menor número de vinculaciones integrales para cada uno de los estados excepto Tamaulipas.

**Gráfica 7 Frecuencias de proyectos de tipo de vinculación y estados**



En la siguiente gráfica se presentan los perfiles por renglón de los estados de la república según tipo de vinculación.

Gráfica 8 Perfiles de estado



En la gráfica 8 se presentan los perfiles por estado. Aquí se observa un comportamiento sin tendencias. Los perfiles más altos para cada una de las categorías Integral, Modular, Puntual y Sin vinculación son Zacatecas, Michoacán, Veracruz y Guerrero respectivamente.

## Análisis de correspondencias

### Área tecnológica y el tipo de vinculación

Antes de realizar el análisis de correspondencias se eliminaron dos categorías de la variable área de conocimiento, a saber, *sin área* y *matemáticas aplicadas y modelación* (MAM). Como se puede observar en la tabla 8, treinta proyectos no tienen área PECITI debido a que las empresas no llenaron dicho campo y no se les pudo asignar alguna otra. Además, sólo hay dos proyectos en MAM y esto genera dificultades en la interpretación gráfica de los perfiles una vez realizado el análisis de correspondencia. De tal manera la tabla que se procesa es la siguiente:

**Tabla 11 Área Peciti y tipo de vinculación modificada**

	Sin				Total general
	Vinculación	Puntual	Modular	Integral	
Biotecnología	20	34	15	20	89
Medicina	11	13	7	9	40
Energía	9	10	6	5	30
Medio ambiente (MA)	16	14	7	3	40
Tecnologías industriales de fabricación (TIF)	66	66	27	20	179
Materiales	8	9	5	8	30
Nanotecnología	5	5	4	5	19
Tecnologías de la información y las telecomunicaciones (TICS)	60	31	11	10	112
Total general	195	182	82	80	539

El estadístico Ji-cuadrada es de 38.18 con 21 grados de libertad. El valor p de la prueba asociado es de 0.01228, por lo que se puede afirmar que existe evidencia suficiente para rechazar la hipótesis de nulidad que establece la independencia entre estas dos variables.

La inercia total es de  $38.18/539=0.070834$ , es decir, la variación total de los datos en la tabla es considerablemente pequeña por lo que se espera que los renglones y las

columnas se encuentren agrupados en una región cercana al centro del plano (el promedio de los perfiles de renglón). Como se puede apreciar en la siguiente tabla, la inercia es representada en tres dimensiones. Las primeras dos explican el 97.08 % de la varianza por lo que los datos están representados de manera fiable en dos dimensiones.

**Tabla 12 Porcentaje de Inercia explicada**

Inercias principales			
	1	2	3
<b>Valor</b>	0.057352	0.009432	0.00207
<b>Porcentaje</b>	83.76%	13.32%	2.92%

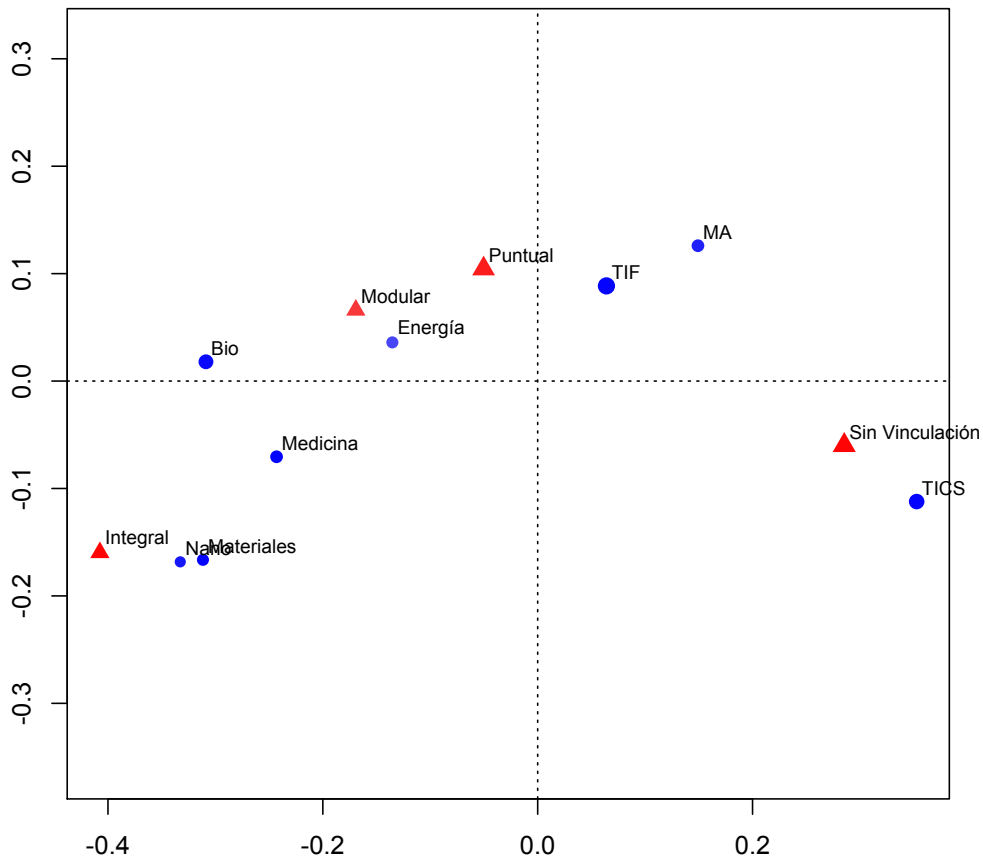
En la gráfica 9, las categorías de las columnas se encuentran a lo largo del eje horizontal. De izquierda a derecha: *integral*, *modular*, *puntual* y *sin vinculación*. Es decir, conforme se avanza en el plano horizontal los proyectos se vuelven más complejos en términos de sus vinculaciones.

Mientras que las categorías *modular* y *puntual* son muy próximas entre sí, las otras (*sin vinculación* e *integral*) son los extremos. De manera cercana a la vinculación modular se encuentran las categorías de *energía* y *medicina*. Los proyectos de TIF parecen estar próximos a la categoría puntual. Medio ambiente parece encontrarse geoméricamente lejana a cualquier perfil de las columnas.

En cuanto al extremo derecho (proyectos sin vinculación) se encuentra el área de *Tecnologías de la información y Telecomunicaciones* (TICS) y en el otro (proyectos con vinculación integral) aparecen las áreas de biotecnología, nanotecnología y materiales.



Gráfica 9 Representación gráfica del AC



En la siguiente tabla se muestra la contribución a la inercia total por renglones. En la primera columna se encuentra la masa, la proporción en decimales y en la segunda en permiles para mayor claridad. Se aprecia que las categorías que mayor contribución a la inercia son TICS y Biotecnología. Como una guía general para definir qué áreas aportaron más a la inercia se usó la sugerencia de Greenacre (2007) de tomar el promedio como umbral. Es decir, si son ocho renglones cada uno debería aportar, en una escala de permiles, 125 como promedio, entonces aquellas categorías que cuenten con un porcentaje más alto se definen como importantes.

**Tabla 13 Masa e inercia por renglón**

	Masa	Inercia	% Permiles
Bio	0.16512	0.016304	<b>230</b>
Energía	0.05566	0.001553	22
MA	0.07421	0.003284	46
Materiales	0.05566	0.006946	98
Medicina	0.07421	0.004767	67
Nano	0.03525	0.005444	77
TICS	0.20779	0.028524	<b>403</b>
TIF	0.33210	0.004008	57
<b>Total</b>	<b>1.000</b>	<b>0.07083</b>	<b>1000</b>

A continuación se muestra la contribución a la inercia de las columnas. De la misma manera se definen las categorías que mayor aportación realizan a la varianza de la tabla. Se puede apreciar que los proyectos que tuvieron una vinculación integral y aquellos que no tuvieron ninguna son los de mayor importancia. Entre ambos contribuyen con más del 80 % de la inercia.

**Tabla 14 Masa e inercia por columna**

	Masa	Inercia	% Permiles
Integral	0.148423	0.028506	<b>402</b>
Modular	0.152134	0.006482	92
Puntual	0.337662	0.005076	72
Sin Vinculación	0.361781	0.030765	<b>434</b>
<b>Total</b>	<b>1.000</b>	<b>0.070829</b>	<b>1000</b>

Por último, se muestra la contribución de cada una de las celdas de la tabla a la inercia total. Una vez más se expresan en una escala de permiles. Si se utilizara la regla del umbral promedio se tendría que todas aquellas celdas por arriba de 31 deberían de definirse como importantes. No obstante, se aprecia que dos celdas dan cuenta de casi el 35 % de la varianza explicada, a saber *Biotecnología-Sin Vinculación* y *TICS-Sin Vinculación*.

**Tabla 15 Contribución de la inercia por celda**

	Sin Vinculación	Puntual	Modular	Integral	Total
Bio	121	14	4	91	230
Energía	8	0	12	2	22
MA	4	0	4	38	46
Materiales	20	3	1	74	98
Medicina	22	0	5	41	68
Nano	13	8	11	45	77
TICS	245	32	56	69	402
TIF	1	13	0	43	57
Promedio	54.25	8.75	11.625	50.375	1000

### Área de conocimiento y grado académico

Antes de realizar el análisis de correspondencias se eliminaron las categorías de sin área de conocimiento y MAM que se señalaron en el ejercicio de AC anterior, y aquellos proyectos que no usaron recursos externos. El motivo de eliminar la categoría *sin recursos externos* es que complicaba el análisis y la representación gráfica de los perfiles. Además, no es de interés conocer el área de aquellos proyectos cuya demanda de conocimiento está satisfecha de manera endógena.

De tal manera la tabla que se procesa para este análisis es la siguiente:

**Tabla 16 Área Peciti y grado académico modificada**

	Sin personal especializado	Al menos 1 Maestro	Al menos 1 Doctor	Total general
Biología	6	9	45	60
Medicina	8	3	10	21
Energía	5	2	9	16
MA	5	2	9	16
TIF	33	22	38	93
Materiales	1	2	7	10
Nano	0	2	9	11
TICS	18	20	16	54
Total general	76	62	143	281

El estadístico Ji-cuadrada tuvo un valor de 40.34 con 14 grados de libertad. El valor p de la prueba es de 0.002261 por lo que se puede afirmar que existe evidencia suficiente para rechazar la hipótesis de nulidad que establece la independencia entre estas dos variables.

La inercia total es mayor que en el ejercicio AC anterior. El valor de la IT es de  $40.34/281 = .14355$ . Como se había mencionado, en los casos en que la inercia es alta la asociación que existe entre los perfiles de los renglones y las columnas es mayor.

Como se puede apreciar en la siguiente tabla, la inercia es representada por dos inercias principales. Éstas explican el 100 % de la varianza por lo que los datos serían representados de manera fiable en dos dimensiones.

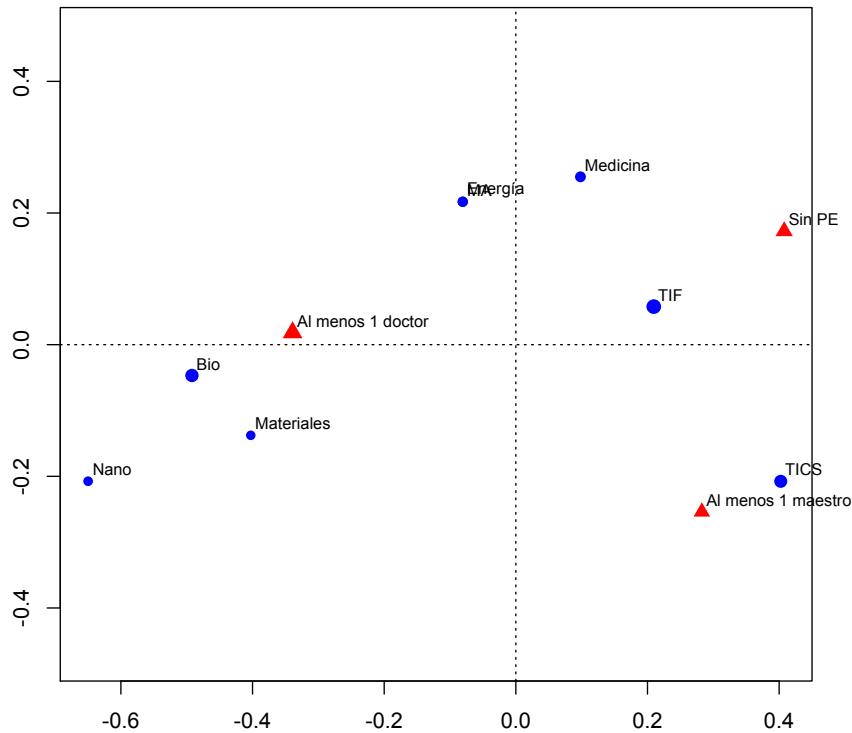
**Tabla 17 Porcentaje de inercia explicada**

Inercias Principales		
	1	2
<b>Valor</b>	0.121102	0.022442
<b>Porcentaje</b>	84.37%	15.63%

En la gráfica 10 se muestra el biplot del AC. El nivel académico de los recursos contratados va aumentando conforme *nos movemos* sobre el eje horizontal y hacia la izquierda. En el extremo derecho se encuentra la categoría de proyectos que contrataron solamente licenciados, ingenieros, personal con especialidad y/o diplomados. Muy cercana a esta categoría se encuentra *al menos 1 maestro*. En el otro extremo se observan los proyectos que requirieron de los recursos más calificados. En este sentido, observamos un eje de las x dividido en dos. En el lado izquierdo se encuentran las disciplinas de biotecnología, nanotecnología, materiales, energía y medio ambiente. Cabe mencionar que las primeras tres se encuentran a la derecha y por debajo del perfil de *al menos 1 doctor*, y las últimas dos yacen a la izquierda y por arriba.

En el lado derecho se encuentra en el cuadrante inferior los proyectos de TICS muy cercanos a la contratación de al menos 1 maestro y por último TIF y Medicina están en el cuadrante superior donde yacen también aquellas empresas que no contrataron a personal especializado.

**Gráfica 10 Representación gráfica del AC**



En la siguiente tabla se muestra la masa, la contribución a la inercia total por renglones en decimales y en porcentaje en permiles. Se aprecia que las categorías que mayor contribución a la inercia son *biotecnología*, *TICS* y *nano*.

**Tabla 18 Masa e inercia por renglón**

	Masa	Inercia	% Permiles
Bio	0.21352	0.05216	<b>363</b>
Energía	0.05694	0.003054	21
MA	0.05694	0.003054	21
Materiales	0.03559	0.006446	45
Medicina	0.07473	0.00558	39
Nano	0.03915	0.018213	<b>127</b>
TICS	0.19217	0.039403	<b>275</b>
TIF	0.33096	0.015634	109
<b>Total</b>	<b>1.000</b>	<b>0.143544</b>	<b>1000</b>

A continuación se muestra la contribución a la inercia de las columnas. De la misma manera se definen las categorías que mayor aportación realizan a la varianza de la tabla. Se puede apreciar que los proyectos que contrataron al menos 1 doctor y que no tuvieron personal especializado aportan una inercia significativa. Entre ambos contribuyen con más del 80 % de la inercia.

**Tabla 19 Masa e inercia por columna**

	Masa	Inercia	% Permiles
Al menos 1 doctor	0.508897	0.058713	<b>409</b>
Al menos 1 maestro	0.220641	0.031842	222
Sin PE	0.270463	0.052989	<b>369</b>
<b>Total</b>	<b>1.000</b>	<b>0.143544</b>	<b>1000</b>

Por último, se muestra la contribución de cada una de las celdas de la tabla a la inercia total. Una vez más se expresan en una escala de permiles. Si se utilizara la regla del umbral promedio se tendría que todas aquellas celdas por arriba de 41 deberían de definirse como importantes. No obstante, se aprecia que cuatro celdas dan cuenta de casi el 60 % de la varianza explicada, a saber *biología- al menos 1 doctor*, *TICS - al menos 1 maestro*, *biotecnología-sin personal especializado* y *TICS - sin personal especializado*.

**Tabla 20 Contribución a la inercia por celda**

	Al menos 1 doctor	Al menos 1 maestro	Sin PE	Total
Bio	<b>160</b>	34	<b>169</b>	363
Energía	3	16	2	21
MA	3	16	2	21
Materiales	27	0	18	45
Medicina	23	14	1	38
Nano	74	2	51	127
TICS	20	<b>136</b>	<b>119</b>	275
TIF	61	3	46	110
Promedio	46.375	27.625	51	1000

### Estado y tipo de vinculación

Otra relación estadísticamente significativa es entre el tipo de vinculación de los proyectos y los estados de la república. En este sentido se pretende explorar las regiones en las que se han realizado un mayor número de proyectos con vinculaciones integrales y modulares. Para este análisis se utilizó la tabla 10 sin modificaciones.

El estadístico Ji-cuadrada presentó un valor de 96.86 con 75 grados de libertad. El valor p de la prueba es de 0.04561 por lo que se puede afirmar que, con un 5 % de nivel de significancia, existe evidencia suficiente para rechazar la hipótesis de nulidad que establece la independencia entre estas dos variables.

La inercia total es de  $96.86/571 = .1696$ . La varianza de los datos en la tabla es considerablemente mayor que en los anteriores AC. En este caso, existe una mayor asociación entre los perfiles de los renglones y las columnas.

Como se puede apreciar en la siguiente tabla, la inercia es representada por tres inercias principales. Las primeras dos explican el 80 % de la varianza por lo que los datos se representan solo de manera aproximada en un plano de dos dimensiones.

**Tabla 21 Porcentaje de inercia explicada**

<b>Inercias principales</b>			
	<b>1</b>	<b>2</b>	<b>3</b>
<b>Valor</b>	0.078717	0.056882	0.034025
<b>Porcentaje</b>	46 %	34 %	20 %

En la gráfica 11 se observa a lo largo del plano las categorías del tipo de vinculación. Una vez más éstas van del extremo derecho al izquierdo aumentando en grado de complejidad. No obstante, es difícil observarlos debido a que hay muy poco espacio entre los perfiles. Cabe mencionar que esto también puede deberse a que algunos puntos podrían ocupar un espacio más alejado si se representaran en 3 dimensiones. Aunque esto hace su interpretación difícil, se puede explicar el comportamiento si observamos 4 grupos muy cercanos.

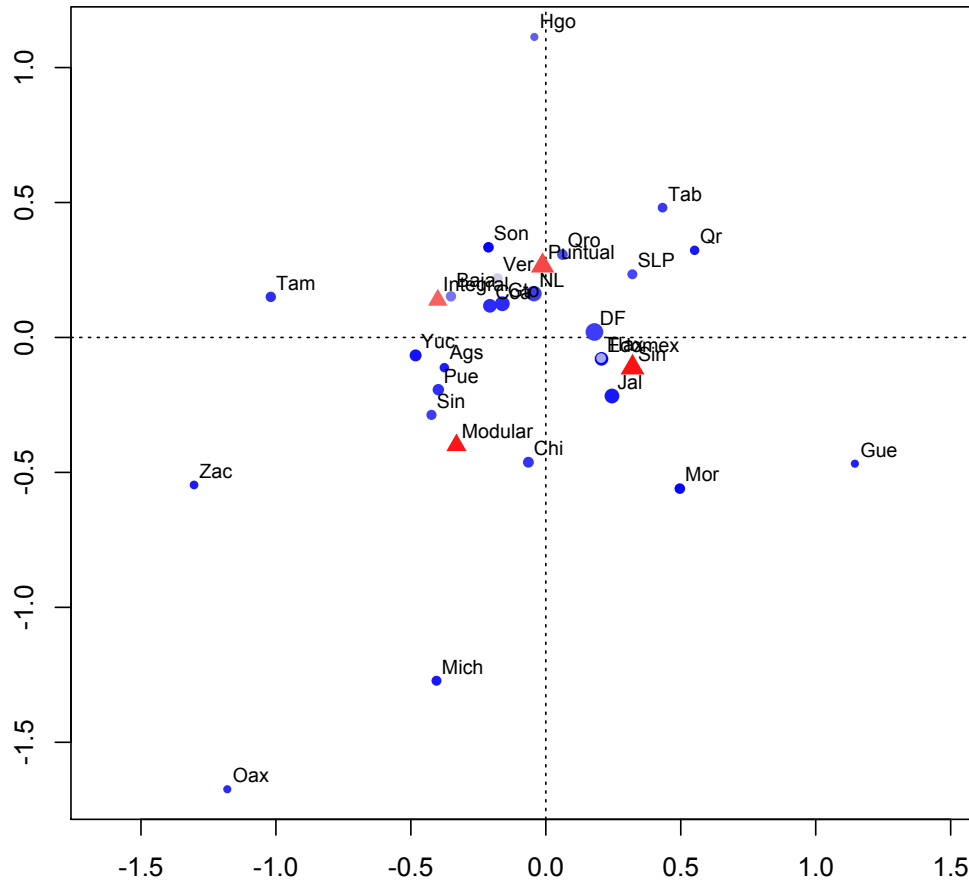
En un primer conjunto se aprecian aquellos estados que se asocian con proyectos que no tuvieron vinculación, como lo son Jalisco, Distrito Federal y Estado de México. En ese cuadrante se encuentran Morelos y Guerrero pero muy alejados del grupo.

En un segundo grupo están los estados más próximos a proyectos con vinculación modular. Aquí se encuentran Yucatán, Aguascalientes, Puebla, Sinaloa, Chihuahua. En este cuadrante se hallan Zacatecas, Michoacán y Oaxaca muy alejados del resto.

Finalmente se aprecian las categorías de Integral y Puntual en un mismo cuadrante, aunque con una menor contribución relativa, cercanos a estas categorías se encuentran los estados de Sonora, Baja California, Nuevo León, Guanajuato, Coahuila y Veracruz. En un cuadrante aparte yacen Querétaro, Quintana Roo, Tabasco y San Luis Potosí.



Gráfica 11 Representación gráfica del AC



En la siguiente tabla se muestra la contribución a la inercia total por renglones en decimales y en porcentaje en permiles. Si se usa la regla del umbral promedio todas aquellas contribuciones mayores a 38 se tomarían como significativas. Se aprecia que el Distrito Federal y los estados de Jalisco, Michoacán, Morelos, Oaxaca, Tamaulipas, Veracruz y Yucatán hacen contribuciones a la inercia mayores que el promedio esperado.

**Tabla 22 Masa e inercia por renglón**

	Masa	Inercia	% Perfiles
Ags	0.007005	0.00122	7
Baja	0.014011	0.004076	24
Chi	0.019264	0.005248	31
Coa	0.073555	0.005011	30
DF	0.255692	0.011221	<b>66</b>
Edomex	0.071804	0.003515	21
Gto	0.108581	0.005486	32
Gue	0.001751	0.00308	18
Hgo	0.001751	0.003596	21
Jal	0.108581	0.01269	<b>75</b>
Mich	0.010508	0.020783	<b>123</b>
Mor	0.015762	0.008834	52
NL	0.136602	0.005125	30
Oax	0.001751	0.008775	<b>52</b>
Pue	0.02627	0.006328	37
Qr	0.007005	0.003173	19
Qro	0.017513	0.002476	15
Sin	0.012259	0.004249	25
SLP	0.010508	0.002336	14
Son	0.015762	0.002482	15
Tab	0.008757	0.004734	28
Tam	0.014011	0.018317	<b>108</b>
Tlax	0.008757	0.001492	9
Ver	0.014011	0.008568	<b>51</b>
Yuc	0.035026	0.008953	<b>53</b>
Zac	0.003503	0.007858	<b>46</b>
<b>Total</b>	<b>1</b>	<b>0.169626</b>	<b>1000</b>

A continuación se muestra la contribución a la inercia de las columnas. Los proyectos que tuvieron una vinculación integral, modular y sin vinculación son los de mayor importancia. Entre ambos contribuyen con más del 80 % de la inercia.

**Tabla 23 Masa e inercia por columna**

	Inercia	% Perfiles
Integral	0.043309	<b>255</b>
Modular	0.049377	<b>291</b>
Puntual	0.032189	190
Sin Vinculación	0.04475	<b>264</b>
Total	0.169625	1000

Finalmente, se muestra la contribución de cada una de las celdas de la tabla a la inercia total. Una vez más se expresan en una escala de permiles. Si se utilizara la regla del umbral promedio se tendría que todas aquellas celdas por arriba de 10 deberían definirse como contribuciones significativas. No obstante, si se analizan las mayores aportaciones de los estados según el tipo de vinculación se observa que aquellos proyectos que no tuvieron actividades de vinculación pertenecen a empresas del DF, Tamaulipas, Jalisco, Yucatán y Morelos principalmente. Es decir, a pesar de que dichos estados contribuyen a la inercia total de manera significativa dicha contribución es en proyectos no vinculados.

Como se mencionó en el análisis de la tabla anterior la categoría de vinculaciones puntuales tiene muy poca participación a la IT. Esto se confirma al observar que la inercia más alta es de 22 del estado de Veracruz. En la siguiente categoría, *vinculación modular*, los estados de Oaxaca y Michoacán tienen las participaciones significativas. Para las vinculaciones integrales son los estados de Tamaulipas y Jalisco los que denotan mayor inercia. No obstante, en un segundo orden de importancia se encuentran Zacatecas, Puebla, Nuevo León y Morelos.

**Tabla 24 Contribución a la inercia por celda**

	Sin vinculación	Puntual	Modular	Integral	Total
Ags	1	1	2	3	7
BC	13	8	3	0	24
Chih	3	19	8	1	31
Coah	19	8	1	2	30
DF	<b>33</b>	3	<b>30</b>	0	66
Edomex	12	2	1	6	21
Gro	12	3	2	1	18
Gto	19	11	1	1	32
Hgo	4	14	2	1	21
Jal	<b>26</b>	6	3	<b>40</b>	75
Mich	0	20	<b>94</b>	9	123
Mor	<b>24</b>	13	2	<b>13</b>	52
NL	2	2	13	<b>13</b>	30
Oax	4	3	<b>44</b>	1	52
Pue	4	8	9	<b>16</b>	37
QR	2	4	7	6	19
Qro	1	9	3	2	15
Sin	1	8	6	10	25
SLP	3	0	10	1	14
Son	5	4	2	4	15
Tab	0	12	9	7	28
Tamps	30	2	3	<b>73</b>	108
Tlax	0	1	0	7	8
Ver	13	<b>22</b>	3	12	50
Yuc	26	0	<b>22</b>	5	53
Zac	7	7	14	<b>18</b>	46
<b>Total</b>	<b>264</b>	<b>190</b>	<b>294</b>	<b>252</b>	<b>1000</b>

## Conclusiones

Como resultado de la prueba Ji-cuadrada aplicada a las variables se concluye que las relaciones estadísticamente significativas son sólo tres. El tipo de vinculación y el grado académico de los recursos humanos externos está asociado al área de conocimiento de los proyectos. La zona geográfica se encuentra relacionada solamente al ejercicio de la vinculación y no al de los recursos humanos o al área de conocimiento de los proyectos.

La complejidad de las actividades de vinculación se encuentra relacionada al conocimiento que subyace en los proyectos. Como era de esperarse aquellas áreas denominadas como tecnológicamente emergentes se hallan asociadas a grados de mayor complejidad. Uno de los resultados más sobresalientes del ejercicio de AC es que el área de conocimiento que domina en esta relación es la de la biotecnología. No obstante, un hallazgo interesante es que esta área también explica gran parte de la asociación con actividades de menor complejidad. Esto se debe, muy probablemente, a la diferenciación del ciclo de la tecnología de las áreas de conocimiento emergentes. Es decir, mientras que el impacto de la nanotecnología, los nuevos materiales, los dispositivos e instrumentación biomédica y la energía alterna ha sido menor en los programas académicos del país, la biotecnología tiene un mayor número de líneas generales de conocimiento asociadas y por lo tanto la oferta en este campo de la ciencia pudiera cubrir las necesidades endógenas de los recursos humanos especializados.

Por otro lado, las tecnologías de información y telecomunicaciones se encuentran asociadas a vinculaciones puntuales. Si bien es un área de conocimiento clave en la economía moderna los proyectos relacionados a TICS en estos fondos no necesitan de vincularse a ninguna institución generadora de conocimiento.

En cuanto al grado académico de los recursos humanos contratados una vez más el área de biotecnología es la que contribuye de mayor manera a la relación entre estas dos variables, asociándose mayormente con aquellos proyectos con el más alto grado académico de contratación. La nanotecnología se encuentra en segundo lugar.

Como vinculación y el personal especializado son actividades estadísticamente independientes no se esperaría observar algún patrón de coincidencia con respecto al área de conocimiento. De manera general, si se cruzan los resultados anteriores en una tabla de dos columnas y dos renglones, que conceptualmente esté dividida por el grado de complejidad en la vinculación y el grado académico podríamos observar que algunas áreas conservan las asociaciones más altas en ambas variables mientras que otras discrepan.

En la tabla 25 se muestra dicha taxonomía (hecha a partir de las inercias obtenidas en cada uno de los ejercicios). Como era de esperarse, áreas como biotecnología, nanotecnología y materiales utilizan recursos humanos más calificados y realizan proyectos de un alto grado de complejidad. Por otro lado, áreas como medicina, medio ambiente y energía no requieren de grados académicos tan altos a pesar de que desarrollan proyectos altamente complejos.

Es decir, podría suceder que las empresas que desarrollan proyectos en estas áreas cuentan con recursos humanos especializados de manera interna o la complejidad de los mismos es limitada a aspectos tecnológicos y no científicos o bien los aspectos de mayor exigencia son realizados por las instituciones con las que se vinculan.

**Tabla 25 Patrones de comportamiento entre el grado de complejidad de vinculaciones y grado académico**

	<b>Nivel grado académico (+)</b>	<b>Nivel grado académico (-)</b>
<b>Complejidad en vinculación (+)</b>	Biotecnología	Medicina
	Nanotecnología	Medio ambiente
	Materiales	Energía
<b>Complejidad en vinculación (-)</b>	TICS	TIF

Sobre la asociación entre los estados de los proyectos y la vinculación se concluye lo siguiente: las necesidades más importantes para la realización de los proyectos se cubren con base en una lógica territorial, las necesidades de recursos humanos de manera endógena, local o contingente.

### **Bibliografía**

- Greenacre, M., (2007), “Correspondence Analysis in Practice” Second edition, Ed. Chapman & Hall, USA.
- Joseph F. Hair Jr. y otros, (2005), “Análisis Multivariante” Quinta edición, Ed. Pearson Prentice Hall, USA.