

| | |
|---|----------|
| 0. INTRODUCCIÓN..... | 1 |
| 1. ANÁLISIS EN COMPONENTES PRINCIPALES (ACP)..... | 4 |
| 1.1. Dominio de aplicación..... | 4 |
| 1.2. Fundamentos del método..... | 4 |
| 1.2.1. Nube de individuos..... | 6 |
| 1.2.2. Solución del ACP, mejor plano de proyección..... | 9 |
| 1.2.3. La nube de variables..... | 11 |
| 1.2.4. ACP normado..... | 14 |
| 1.2.5. Elementos suplementarios..... | 17 |
| 1.3. Un ejemplo de aplicación: perfil productivo de los investigadores..... | 20 |
| 1.3.1. Presentación..... | 20 |
| 1.3.2. Análisis de tablas y gráficos..... | 20 |
| 1.3.3. Conclusiones..... | 23 |
| 1.4. Ejercicio: test de personalidad..... | 24 |
| 1.4.1. Presentación..... | 24 |
| 1.4.2. Objetivos..... | 25 |
| 1.4.3. Tablas y gráficos..... | 25 |
| 1.4.4. Guía para el análisis..... | 29 |

TABLAS Y GRÁFICOS

| | |
|--|----|
| Gráfico 0-1: Esquema de los métodos factoriales y de clasificación..... | 3 |
| Tabla 1-1: Calificación de 10 muestras de café..... | 5 |
| Gráfico 1-1: Diagrama de dispersión Aroma x Cuerpo..... | 6 |
| Tabla 1-2: Matriz X, valores centrados..... | 7 |
| Gráfico 1-2: Nube de puntos centrada en el plano Aroma-Cuerpo..... | 7 |
| Tabla 1-3: Varianza y contribución a la inercia de cada una de las variables..... | 8 |
| Tabla 1-4: Contribución de los individuos a la inercia..... | 8 |
| Gráfico 1-3: Primer plano factorial del ACP aroma-cuerpo..... | 10 |
| Tabla 1-5: Matriz de covarianzas para la Tabla 1-1..... | 10 |
| Tabla 1-6: Histograma de valores propios..... | 10 |
| Gráfico 1-4: Primer plano factorial del ACP no normado de la Tabla 1-1..... | 11 |
| Gráfico 1-5: Significado de la media y del centrado en la nube de variables..... | 12 |
| Tabla 1-7: Coordenadas, correlaciones con los ejes y antiguos ejes unitarios para las variables..... | 14 |
| Gráfico 1-6: Primer plano factorial de las variables del ACP no normado de las muestras de café..... | 13 |
| Gráfico 1-7: Ejemplo de representación de 4 variables y su proyección..... | 14 |

| | |
|---|----|
| Tabla 1-8: Matriz de correlaciones | 15 |
| Tabla 1-9: Histograma de los valores propios del ACP normado “café” | 15 |
| Tabla 1-10: Coordenadas y antiguos ejes unitarios del ACP normado “café” | 16 |
| Tabla 1-11: Coordenadas, contribuciones y cosenos cuadrados de las muestras de café (ACP normado) | 16 |
| Gráfico 1-8: Círculo de correlación para el ejemplo café..... | 16 |
| Gráfico 1-9: Primer plano factorial del ACP normado del ejemplo café | 17 |
| Tabla 1-12: Tabla de datos mostrando los promedios según las modalidades de agregado..... | 18 |
| Gráfico 1-10: Primer plano factorial con individuos y modalidades ilustrativas | 19 |
| Gráfico 1-11: Impresión proyectada como variable continua ilustrativa | 19 |
| Tabla 1-13: Coordenadas y valores test de las modalidades | 18 |
| Tabla 1-14: Estadísticas básicas de las variables continuas del ACP..... | 20 |
| Tabla 1-15: Matriz de correlaciones y valores test para la producción de los investigadores..... | 20 |
| Tabla 1-16: Valores propios del ACP e histograma | 21 |
| Tabla 1-17: Coordenadas y antiguos ejes unitarios para las variables continuas | 22 |
| Gráfico 1-12: Círculo de correlaciones (primer plano factorial de las variables del ACP)..... | 22 |
| Gráfico 1-13: Primer plano factorial con modalidades ilustrativas | 23 |
| Tabla 1-18: Coordenadas y valores test de las variables nominales ilustrativas | 23 |
| Gráfico 1-14: investigadores sobre el primer plano factorial | 24 |
| Tabla 1-19: Fragmento.de.la.Base.de.Datos:.Puntajes.en.laPrueba.Sicológica.PPG-IPG y en la Prueba 16PF..... | 25 |
| Tabla 1-20: Estadísticas básicas de las variables continuas del ACP..... | 26 |
| Tabla 1-21: matriz de correlaciones | 26 |
| Tabla 1-22: Valores propios del ACP e histograma | 27 |
| Tabla 1-23: Coordenadas y ejes unitarios para las variables continuas..... | 27 |
| Gráfico 1-15: Primer Plano Factorial. Proyección de los Individuos | 28 |
| Gráfico 1-16: Primer Plano Factorial. Proyección de la Variables Activas | 28 |
| Gráfico 1-17: Primer plano factorial. proyección de las variables activas, conjuntamente con las Ilustrativas | 29 |

MÉTODOS ESTADÍSTICOS MULTIVARIADOS EN INVESTIGACIÓN SOCIAL

PRINCIPIOS Y EJEMPLOS DE APLICACIÓN

0. INTRODUCCIÓN

En Ciencias Humanas y Sociales se suelen denominar métodos cuantitativos, en contraposición a los cualitativos, a aquellos que buscan documentar con datos numéricos los fenómenos que están siendo observados sobre poblaciones, muestras o grupos mas o menos grandes. Es posible conseguir indicadores de escala continua, pero la información original es generalmente de escala nominal u ordinal. De cualquier manera la información sobre las unidades de observación ("individuos") se transforma en tablas de datos. Una tabla de datos generalmente tiene filas que representan a los "individuos" y columnas que representan a las variables, las cuales pueden ser continuas o nominales según la escala de medición.

Lo que interesa a los investigadores, en primera instancia, es hacer una lectura de la información contenida en la tabla de datos y entonces son pertinentes las siguientes preguntas:

- ¿Cómo abordar la información que hay en una tabla de datos?
- ¿Cómo leer una tabla de datos?, es decir, ¿qué información importante hay en una tabla de datos en relación con los objetivos del estudio?.
- ¿Cómo obtener un mensaje que pueda ser luego verbalizado por el investigador y sirva para la comunicación de los resultados?

Para lograr que el cerebro humano pueda captar lo más importante de la información de una tabla hay que consentir perder información para ganar en significación. El cerebro entiende mejor la información en forma análoga, es decir en forma gráfica, en lugar de la información digital o el conjunto de cifras de una tabla. Al menos en un comienzo es de gran ayuda observar gráficas que representen de alguna manera la información más importante de las cifras y símbolos puestos en la tabla. Uno de los elementos de la estadística descriptiva que cumple con ese cometido es el histograma de frecuencias de una variable cuantitativa.

La naturaleza de la información en Ciencias Humanas y Sociales hace que casi la única alternativa para el análisis de la información de sus investigaciones, cuando es posible presentarla en tablas de datos, es recurrir a los métodos estadísticos descriptivos y exploratorios, de los cuales, los más importantes, se pueden agrupar así:

- Estadística descriptiva univariada y bivariada
- Análisis exploratorio de datos (EDA)
- Métodos estadísticos exploratorios multidimensionales.

Los métodos descriptivos univariados y bivariados forman parte de los cursos y textos de estadística o métodos cuantitativos de los programas académicos de las diferentes disciplinas. Algunos de los métodos de EDA se han venido incluyendo también en los textos básicos de estadística. Estos dos primeros tipos de métodos a lo sumo tienen en cuenta la relación entre dos variables y de alguna manera se pueden extender a tres variables, pero es el tercer grupo de métodos el que hace posible tener en cuenta la interrelación entre múltiples variables. Los investigadores saben que la información de los fenómenos sociales es esencialmente de naturaleza multivariada. Es por eso que el objetivo de este texto es hacer una

introducción al lenguaje, lógica y aplicación de los métodos, para que los investigadores puedan recurrir a ellos.

Lebart et al (1995) han acuñado para estos métodos el nombre de exploratorios multidimensionales, pero se usó mucho en el pasado el de *análisis de datos* y es sinónimo de *estadística descriptiva multivariada* o *análisis multivariado de datos*. Se constituyen en una generalización de la estadística descriptiva univariada y bivariada, pero la presencia de más variables o dimensiones la hace más compleja. La interpretación de las representaciones gráficas requieren del conocimiento de la lógica de los métodos y están siempre acompañadas de índices numéricos que complementan y enriquecen los análisis. En otras palabras la utilización de estos métodos requiere de un entrenamiento para su utilización e interpretación y hace prácticamente indispensable el trabajo interdisciplinario en la investigación. Solamente es posible la utilización adecuada de estos métodos por una sola persona cuando ella juega el doble papel de investigador social y estadístico

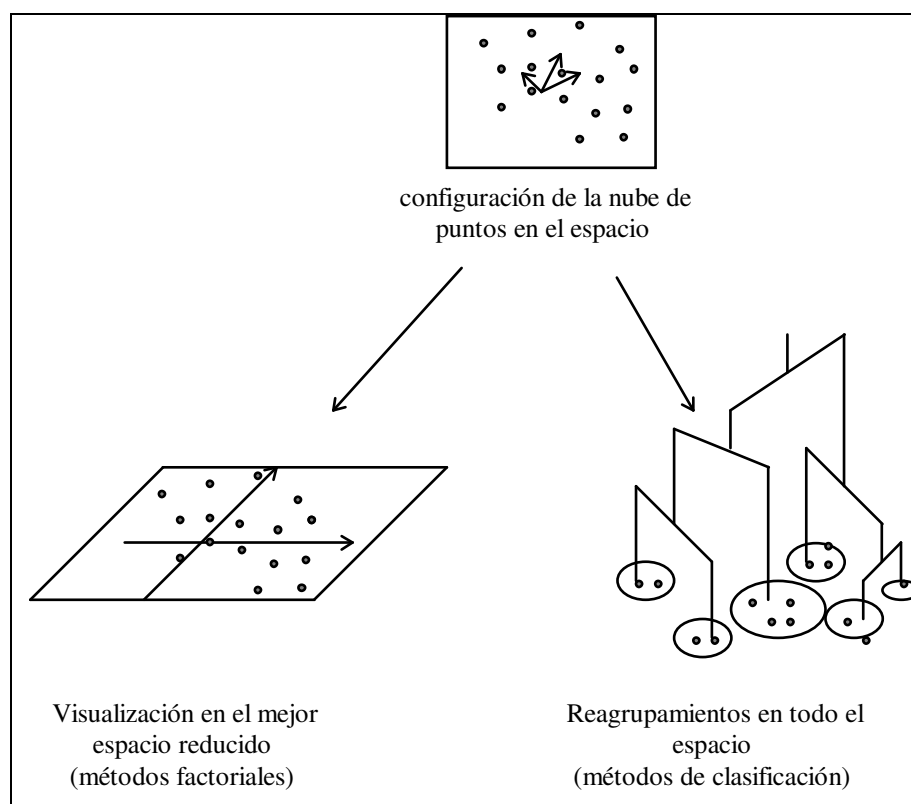
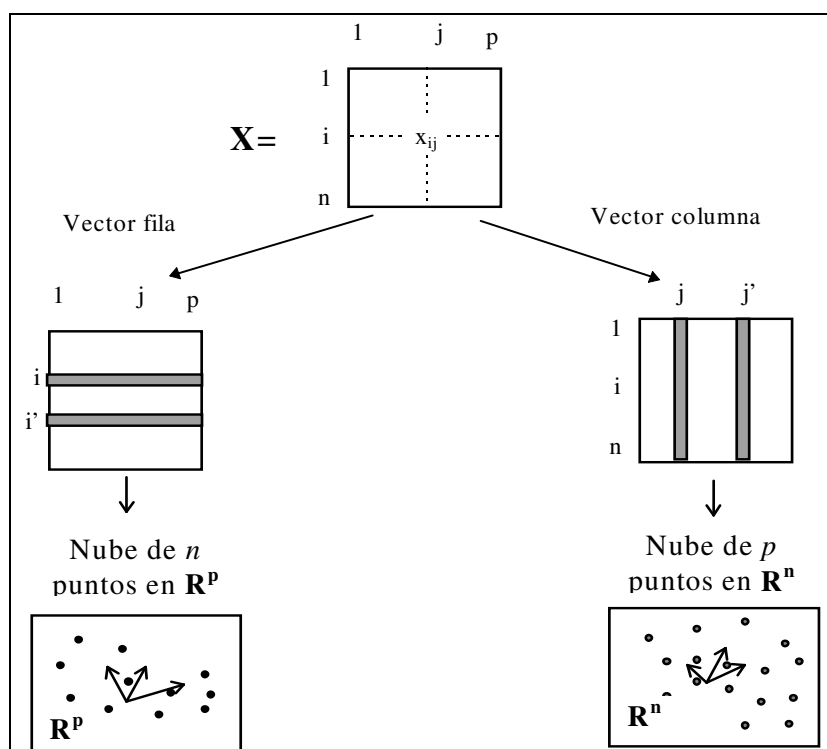
Siendo el objetivo de estos métodos la descripción y exploración de la información no se requiere de modelos preestablecidos, ni de supuestos que muchas veces no se cumplen. Los métodos logran la presentación analógica de la información recurriendo a principios geométricos. La tabla de datos se representa, luego de una transformación adecuada, en un espacio de múltiples dimensiones: *nube de puntos*. En la representación geométrica la distancia entre puntos significa la diferencia entre los elementos considerados: si están cerca se parecen, si están lejos son muy diferentes.

La nubes de puntos construidas son abstractas pues no podemos ver espacios de más de tres dimensiones, en realidad, en nuestros documentos, vemos bien dos dimensiones (planos). Pero la geometría abstracta de esas representaciones multidimensionales cumple con las mismas propiedades de la geometría plana y del espacio. Se recurre entonces a proyecciones sobre planos y a agrupamientos de puntos cercanos, para observar lo más importante de esas representaciones multidimensionales. La lectura utilizando proyecciones es el principio de los *métodos factoriales*, en cuyo caso la pérdida de la información se manifiesta en forma de errores de proyección. En los métodos factoriales se busca el plano para el cual los errores de proyección son en conjunto los menores posibles: *primer plano factorial*.

La lectura de la representación mutidimensional conformando clases o grupos de puntos cercanos, constituye los *métodos de clasificación*. En estos métodos la pérdida de información se da porque cada elemento pierde sus características específicas y se caracteriza, en cambio, por la clase a la que pertenece. Se buscan grupos de tal manera que los elementos al interior de un grupo se parezcan y los elementos de diferentes grupos sean lo más diferenciados posible. El Gráfico 0-1 (tomada de Lebart et al. (1995)), ilustra el principio geométrico de los métodos factoriales y de clasificación. La nomenclatura que aparece es el gráfico es propia del álgebra lineal. Sobre una tabla de datos son posibles dos representaciones complementarias: *la nube de los puntos fila* y *la nube de los puntos columna*. Para ubicar un punto en el plano se requieren dos coordenadas, para ubicarlo en el espacio 3 coordenadas y para ubicarlo en un espacio abstracto de p dimensiones, p coordenadas. El conjunto de las coordenadas necesarias para ubicar un punto se denomina *vector*. En una tabla de n filas y p columnas, se tiene una nube de n puntos filas en donde cada fila está representada mediante un vector de p coordenadas y una nube de p puntos columna con cada punto representado por un vector de n coordenadas.

La naturaleza de las filas y columnas de una tabla de datos junto con los objetivos del estudio determinan los métodos a utilizar. Dentro de los métodos factoriales el más útil en ciencias humanas y sociales es el análisis de correspondencias múltiples (ACM), ya que es el adecuado para la lectura de tablas de “individuos” por variables cualitativas (nominales u ordinales). El ACM es una generalización del análisis de correspondencias simples (ACS), utilizado para la lectura de tablas de contingencia. El ACS se puede ver como la aplicación simultánea de dos análisis en componentes principales (ACP). Aún bajo el supuesto de que un investigador este interesado en la aplicación del ACM es necesario conocer previamente los otros dos métodos. En la mayoría de las aplicaciones se utilizan métodos de clasificación que dan lecturas complementarias a los métodos factoriales, de la tabla de datos.

Gráfico 0-1: Esquema de los métodos factoriales y de clasificación



Este texto se estructura en un capítulo por cada uno de los métodos mencionados. Los ejemplos están realizadas en distintas versiones del programa SPAD de origen francés (www.cisia.com). El propósito es hacer una presentación comprensible para investigadores de diferentes áreas. Sin embargo no es posible una adecuada comprensión de la lógica de los métodos sin un entendimiento de la geometría al menos desde un punto de vista intuitivo. Los interesados en conocer mejor los procedimientos matemáticos pueden recurrir a algunas de las referencias dadas.

1. ANÁLISIS EN COMPONENTES PRINCIPALES (ACP)

1.1. *Dominio de aplicación*

El análisis en componentes principales es útil en la lectura de tablas de “individuos” por variables cuantitativas. Se tiene por costumbre escribir en las filas a los “individuos” los que representan las unidades estadísticas en un análisis. En el ejemplo de aplicación se tienen 2148 investigadores como individuos, los cuales están descritos por la producción de resultados de diferente tipo: artículos en revistas científicas, libros, capítulos de libro, etc. Para ilustrar la lógica del método se emplea un ejemplo sencillo en donde los individuos son 10 muestras de café y las variables las calificaciones promedio de un panel de catadores según diferentes aspectos. Como ejercicio se utilizan los resultados obtenidos al aplicar un test de personalidad a un grupo de 1142 personas, las variables son los distintos items del test constituidos por puntajes entre 0 y 100.

El objetivo principal del ACP es comparar los individuos según los valores de las variables continuas que se parecen. Dos investigadores serán muy parecidos porque producen más o menos las mismas cantidades de los mismos tipos de resultados. Dos muestras de café se parecen porque obtienen más o menos las mismas calificaciones en los diferentes aspectos.

Un segundo objetivo es el de ver relaciones entre las variables que están describiendo a los individuos y de el se deriva un tercer objetivo, útil cuando se requieren otros análisis posteriores, que es el de reducir la dimensionalidad del problema. Una alta correlación entre las variables dará como resultado que unas pocas variables sintéticas resuman lo importante de la información de las variables originales.

1.2. *Fundamentos del método*

El ACP recurre a dos representaciones geométricas: una para comparar a los individuos (nube de individuos) y otra para estudiar las relaciones entre las variables (nube de variables). Estas representaciones requieren de transformaciones de la tabla de datos. La transformación más utilizada es la de la estandarización de los datos, es decir restar la media (centrado) y dividir por la desviación estándar (reducido), lo que da origen al análisis en componentes principales normado. Pero la comprensión del ACP es más sencilla siguiendo el proceso para el ACP no normado, que corresponde a una transformación de centrado solamente (resta de la media).

Un lenguaje matemático mínimo es indispensable para entender estas ideas. Los números de la Tabla 1-1 corresponden a la matriz de datos que el ACP utilizará, este arreglo de números se denota con la letra **R**, cada celda de la matriz se identifica con una letra con dos subíndices el primero para las filas y el segundo para las columnas. Así r_{ij} representa el valor que toma el individuo i para la variable j , por ejemplo $r_{32} = 5.98$ es la calificación de aroma para la muestra de café claro con 40% de cebada.

Tabla 1-1: Calificación de 10 muestras de café

| IDEN | IntensidadAroma | Aroma | Cuerpo | AcidezTasa | Amargo | Astringencia |
|--------------|-----------------|-------|--------|------------|--------|--------------|
| ExcelsoClaro | 7.72 | 7.00 | 6.84 | 5.02 | 5.04 | 5.36 |
| Claro40Maiz | 6.02 | 5.42 | 6.22 | 4.34 | 4.60 | 4.78 |
| Claro40Cebad | 6.48 | 5.98 | 6.44 | 4.58 | 4.82 | 4.80 |
| Claro20Maiz | 6.82 | 6.44 | 6.70 | 4.62 | 4.38 | 4.80 |
| Claro20Cebad | 7.08 | 6.20 | 6.72 | 4.78 | 4.94 | 4.90 |
| ExcelsoOscur | 7.66 | 7.42 | 6.98 | 5.12 | 5.18 | 5.22 |
| Oscuro40Maiz | 6.18 | 5.82 | 6.26 | 4.00 | 4.46 | 4.96 |
| Oscuro40Ceba | 6.84 | 6.56 | 6.82 | 4.30 | 4.96 | 4.84 |
| Oscuro20Maiz | 6.66 | 7.06 | 6.70 | 4.64 | 5.00 | 4.90 |
| Oscuro20Ceba | 7.00 | 6.70 | 7.04 | 4.60 | 4.88 | 5.18 |

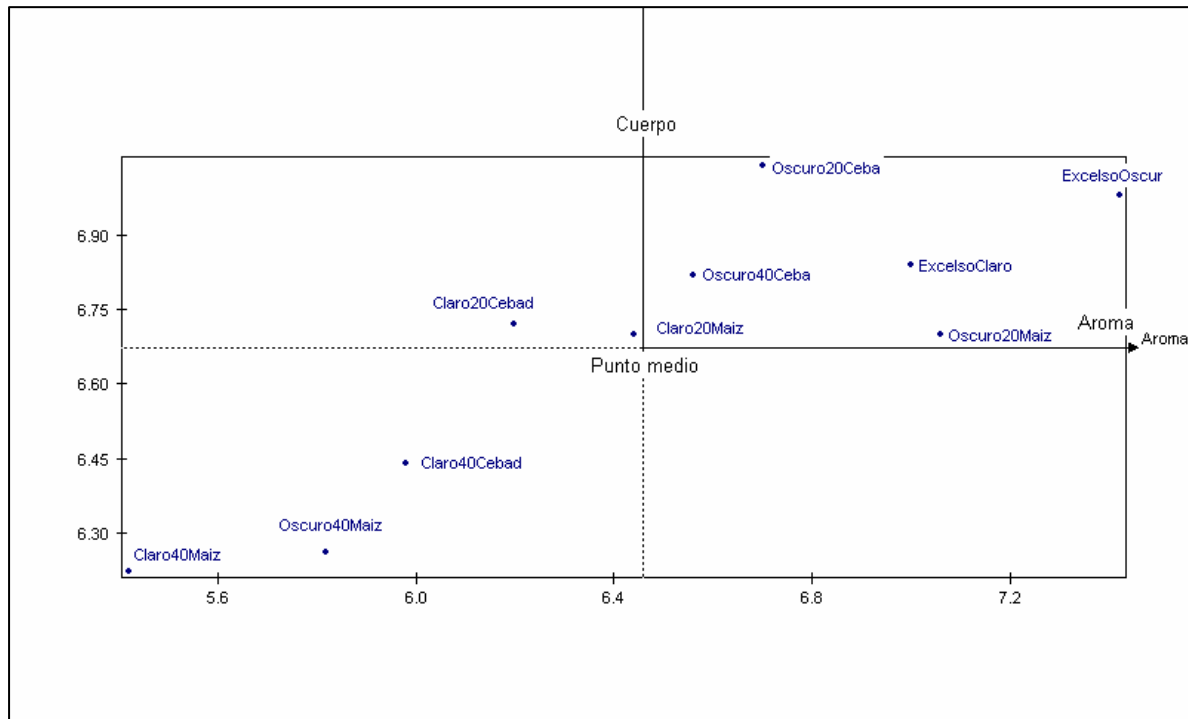
Nota: “Las muestras escogidas fueron: café excelso (*Coffea arabica*), maíz amarillo (*Zea mays*) y cebada perlada (*Hordeum vulgare*) tostados y molidos. El café fue torrefactado a 230°C y se obtuvieron dos muestras: una de tostación clara (8 minutos) y otra de tostación oscura (10 minutos). La cebada y el maíz se tostaron a 230°C por 20 minutos, para lograr una coloración similar a la del café. Además se hicieron análisis en mezcla de café y cebada, y de café y maíz, al 20 y 40%, respectivamente”. (Duarte et. al., 1996; página 66).

La tabla **R** tiene $n = 10$ filas y $p = 6$ columnas, cada fila tiene 6 valores que corresponden a las calificaciones que cada muestra de café obtiene según cada uno de los 6 aspectos. Matemáticamente este arreglo de 6 números se puede ver como un *vector* de 6 dimensiones. Cada columna tiene 10 valores que representan la calificación obtenida por cada una de las muestras según el aspecto (variable) de la columna. Cada arreglo de 10 valores es un vector de 10 dimensiones. En estadística los vectores se suelen escribir en forma de columna y cuando están en fila se toma como el transpuesto.

Llamando \mathbf{w}_i al vector que tiene los valores de las variables para la fila i , se tiene por ejemplo, para la muestra de café claro con 40% de cebada el siguiente vector de calificaciones: $\mathbf{w}'_3 = [6.48, 5.98, 6.44, 4.58, 4.82, 4.80]$. Denotando una variable como \mathbf{x}_j , se tiene por ejemplo para el aroma el siguiente vector de calificaciones para las 10 muestras:

$$\mathbf{x}'_j = [7.00, 5.42, 5.98, 6.44, 6.20, 7.42, 5.82, 6.56, 7.06, 6.70]$$

La tabla de datos **R** acepta entonces dos miradas: 10 vectores fila cada uno de 6 dimensiones y 6 vectores columna cada uno de 10 dimensiones. Los 60 números de la tabla representan la información contenida en ella. Para comparar los individuos se utiliza la representación geométrica de los 10 vectores fila, cada fila se representa como un punto en un sistema de 6 ejes. Cada uno de los 6 números se constituye entonces en la coordenada sobre un eje en una geometría canónica. Esta representación es abstracta pero con todas las propiedades de 2 y 3 dimensiones que nos son familiares. La representación se puede entender fácilmente en dos dimensiones tomando por ejemplo las variables aroma y cuerpo solamente, entonces los 10 vectores son: $[7.00, 6.84]$, $[5.42, 6.22]$, $[5.98, 6.44]$, $[6.44, 6.70]$, $[6.20, 6.72]$, $[7.42, 6.98]$, $[5.82, 6.26]$, $[6.56, 6.82]$, $[7.06, 6.70]$, $[6.70, 7.04]$, (ver Gráfico 1-1). En el gráfico también aparece el punto medio (vector cuyas coordenadas son los promedios de cada una de las variables), el cual se suele denominar centro de gravedad.

Gráfico 1-1: Diagrama de dispersión Aroma x Cuerpo

Partiendo de la representación completa en los 6 ejes, el Gráfico 1-1 es una proyección en el plano Aroma-Cuerpo. El fundamento del ACP es precisamente buscar un plano de proyección que permita ver lo mejor posible los puntos que están representados en las seis dimensiones. La representación de estos vectores fila constituye la nube de individuos y la de los vectores columna la nube de variables.

1.2.1. Nube de individuos

En ACP la representación de los individuos se hace colocando los ejes en el centro de gravedad. Las coordenadas para esta representación se obtienen restando a cada uno de los 10 vectores el vector centro de gravedad \mathbf{g} , cuyas coordenadas son los 6 promedios de cada una de las variables. Tomando los promedios de las columnas de la Tabla 1-1, se obtiene:

$$\mathbf{g}' = [6.85, 6.46, 6.67, 4.60, 4.83, 4.97]$$

Es decir que cada coordenada del centro de gravedad es en términos generales:

$$\bar{r}_i = \frac{1}{n} \sum_{j=1}^n r_{ij}$$

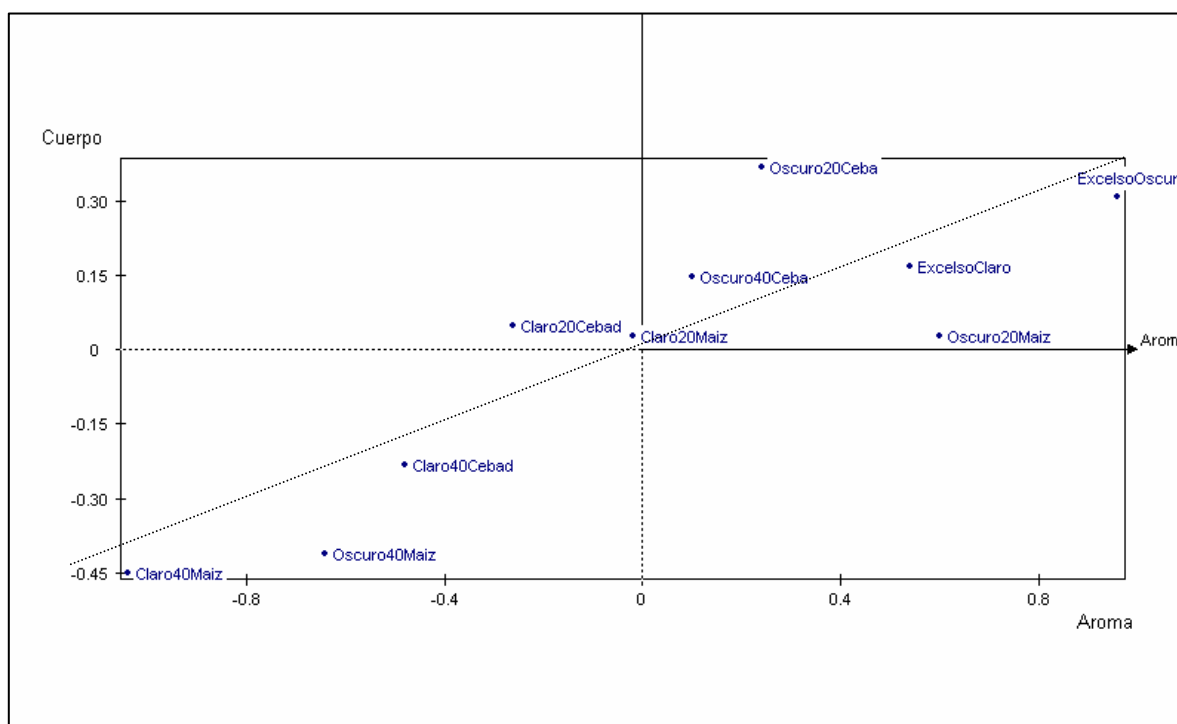
En el Gráfico 1-1 el punto medio es [6.46, 6.67], cuyos números son las segunda y tercera coordenada de \mathbf{g}' , correspondiente a los promedios de aroma y cuerpo. Si se denomina \mathbf{X} (Tabla 1-2), a la tabla centrada, cada elemento de esta tabla se obtiene mediante:

$$x_{ij} = r_{ij} - \bar{r}_j$$

Tabla 1-2: Matriz X, valores centrados

| IDEN | IntensidadAroma | Aroma | Cuerpo | AcidezTasa | Amargo | Astringencia |
|--------------|-----------------|-------|--------|------------|--------|--------------|
| ExcelsoClaro | 0.87 | 0.54 | 0.17 | 0.42 | 0.21 | 0.39 |
| Claro40Maiz | -0.83 | -1.04 | -0.45 | -0.26 | -0.23 | -0.19 |
| Claro40Cebad | -0.37 | -0.48 | -0.23 | -0.02 | -0.01 | -0.17 |
| Claro20Maiz | -0.03 | -0.02 | 0.03 | 0.02 | -0.45 | -0.17 |
| Claro20Cebad | 0.23 | -0.26 | 0.05 | 0.18 | 0.11 | -0.07 |
| ExcelsoOscur | 0.81 | 0.96 | 0.31 | 0.52 | 0.35 | 0.25 |
| Oscuro40Maiz | -0.67 | -0.64 | -0.41 | -0.60 | -0.37 | -0.01 |
| Oscuro40Ceba | -0.01 | 0.10 | 0.15 | -0.30 | 0.13 | -0.13 |
| Oscuro20Maiz | -0.19 | 0.60 | 0.03 | 0.04 | 0.17 | -0.07 |
| Oscuro20Ceba | 0.15 | 0.24 | 0.37 | 0.00 | 0.05 | 0.21 |

La representación geométrica de las 10 filas como 10 puntos en un espacio de 6 dimensiones (\mathbf{R}^6), constituye la nube de individuos centrados y contiene toda la información desde el punto de vista de comparación de los individuos. El centro de la representación significa ahora un individuo promedio, un punto de referencia apropiado para la comparación. El Gráfico 1-2, muestra la nube de puntos centrada, para las variables aroma y cuerpo. Nótese que es la misma imagen del Gráfico 1-1, se diferencian en que las coordenadas de los puntos, en este último gráfico, significan la diferencia con el individuo promedio.

Gráfico 1-2: Nube de puntos centrada en el plano Aroma-Cuerpo

El parecido o diferencia de los individuos según las variables que los están caracterizando se traduce en la analogía geométrica en un sistema de distancias. El ACP utiliza la distancia euclidiana canónica, es decir la que aprendimos desde la geometría elemental: la distancia al cuadrado entre dos puntos es la suma de las diferencias al cuadrado entre las coordenadas, o sea:

$$d^2(i, l) = \sum_{j=1}^p (x_{ij} - x_{lj})^2$$

El parecido o diferencia de los individuos se traduce en cercanía o lejanía geométrica y entonces la lectura de las gráficas permite observar lo más importante de la tabla de datos.

No podemos leer en un documento sino dos dimensiones y entonces necesitamos el mejor plano para aproximarnos a la lectura, es algo así como encontrar la mejor fotografía de la nube de puntos.

Poner el sistema de referencia en el centro de gravedad conlleva olvidarse de la posición de la nube de puntos y concentrarse en su dispersión, es decir la forma de la nube de puntos.

Una medida global de la dispersión se obtiene definiendo la inercia de la nube de puntos, que es el promedio de la suma de las distancias entre cada punto y el centro de gravedad (recuérdese que para la nube centrada las coordenadas del centro de gravedad son todas cero) y resulta ser la suma de las varianzas de todas las variables:

$$I = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p x_{ij}^2 = \sum_{j=1}^p \frac{1}{n} \sum_{i=1}^n x_{ij}^2 = \sum_{j=1}^p V(r_j)$$

El concepto de inercia es central tanto en los métodos factoriales como de clasificación, es la generalización del concepto de varianza al caso de varias variables.

La nube de puntos posee una inercia que se puede descomponer en cada uno de los ejes de la representación, en planos y en subespacios de mayor dimensión. Cada eje contribuye a la inercia con la varianza de la variable que representa, es decir corresponde a la contribución de la variable a la inercia total. En la Tabla 1-3 se ve que la variable aroma contribuye más a la inercia (38%) y la variable cuerpo es una de las que menos contribuye (8%). Esto se refleja en el Gráfico 1-2, con una mayor dispersión de la nube en el eje horizontal (aroma).

Tabla 1-3: Varianza y contribución a la inercia de cada una de las variables

| Variable | IntensidadAroma | Aroma | Cuerpo | AcidezTasa | Amargo | Astringencia | Inercia |
|----------|-----------------|-------|--------|------------|--------|--------------|---------|
| Varianza | 0.28 | 0.34 | 0.07 | 0.10 | 0.06 | 0.04 | 0.89 |
| %Inercia | 31.32 | 38.37 | 7.97 | 11.18 | 6.93 | 4.24 | 100.00 |

Cada individuo contribuye con el cuadrado de su distancia al centro de gravedad dividida por n (total de individuos). Los individuos más alejados del centro tienen más importancia en el análisis (las muestras excelso y 40% maíz en el ejemplo).

Tabla 1-4: Contribución de los individuos a la inercia

| IDEN | Inercia | %Inercia |
|---------------|---------|----------|
| ExcelsoClaro | 0.15 | 16.27 |
| Claro40Maiz | 0.21 | 23.76 |
| Claro40Cebad | 0.04 | 5.02 |
| Claro20Maiz | 0.02 | 2.59 |
| Claro20Cebad | 0.02 | 1.96 |
| ExcelsoOscuro | 0.21 | 23.88 |
| Oscuro40Maiz | 0.15 | 16.97 |
| Oscuro40Ceba | 0.02 | 1.77 |
| Oscuro20Maiz | 0.04 | 4.84 |
| Oscuro20Ceba | 0.03 | 2.93 |
| Total | 0.89 | 100.00 |

1.2.2. Solución del ACP, mejor plano de proyección

Toda la información de interés para la comparación de los individuos está representada geoméricamente en un espacio de muchas dimensiones (\mathbf{R}^p), observar esa representación, de alguna manera, es el objetivo del ACP. Para lograrlo hay que observar las proyecciones de la nube de individuos sobre ejes y planos. En primer lugar se busca la dirección de un primer eje que conserve mejor las distancias originales de los puntos, que es igual a conservar las distancias de los puntos al centro de gravedad. Este es un problema de optimización que se resuelve por el método de mínimos cuadrados y que da como resultado que la dirección de ese eje es la misma del vector propio asociado al valor propio más grande de la matriz $\mathbf{X}'\mathbf{X}$. Luego se busca un segundo eje perpendicular al anterior que recoja lo más posible la dispersión remanente y así sucesivamente. La solución es entonces la diagonalización de la matriz $\mathbf{X}'\mathbf{X}$, es decir obtener sus valores propios ordenados de mayor a menor y los vectores propios asociados. En el ACP no normado la matriz $\mathbf{X}'\mathbf{X}$ que se diagonaliza resulta ser la matriz de varianzas y covarianzas, la cual aparece en las salidas de los programas de ACP. Geométricamente este resultado corresponde a una rotación del sistema de ejes. El primer eje tiene la dirección más alargada de la nube o sea la de mayor dispersión, que corresponde a la dirección de mayor inercia. El plano conformado por los dos primeros nuevos ejes, denominados factoriales, es la mejor fotografía de la nube de puntos.

Para entender el problema, supóngase que solo se tienen las variables aroma y cuerpo, es decir las 10 muestras descritas por las dos variables. La nube de puntos centrados es el Gráfico 1-2, la línea diagonal

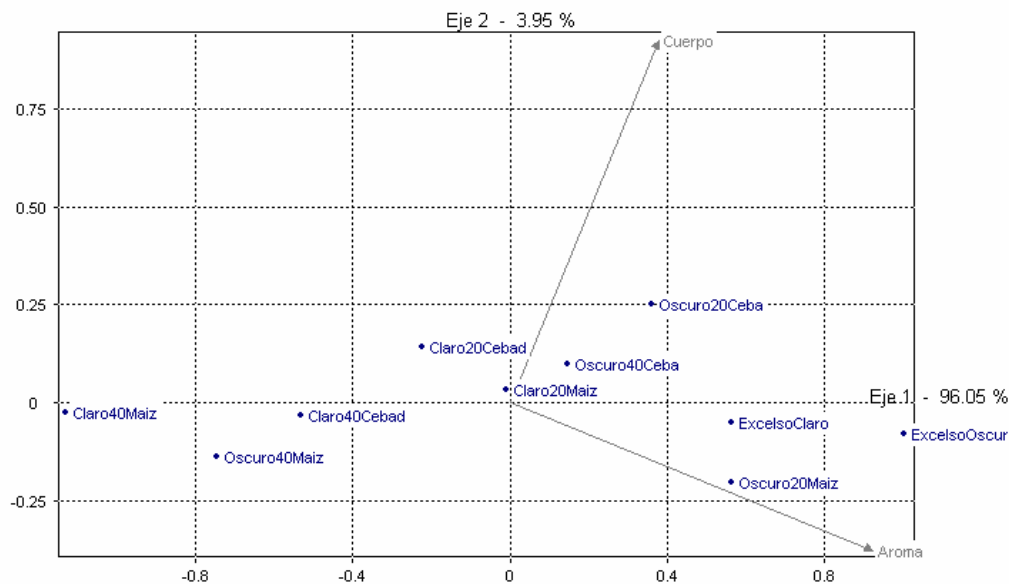
discontinua corresponde al primer eje del ACP. La matriz de covarianzas es: $\mathbf{V} = \begin{bmatrix} 0.34 & 0.13 \\ 0.13 & 0.07 \end{bmatrix}$, la

inercia total es 0.41 (0.34+0.07). Los dos valores propios son: 0.398 y 0.016, los vectores propios son:

$\mathbf{u}_1 = \begin{bmatrix} 0.93 \\ 0.38 \end{bmatrix}$ y $\mathbf{u}_2 = \begin{bmatrix} -0.38 \\ 0.93 \end{bmatrix}$. La inercia se descompone ahora sobre los nuevos ejes y para cada uno es

el valor propio asociado al eje, entonces el primer eje factorial contribuye a la inercia con el 96%. Para construir la gráfica usando los nuevos ejes es necesario encontrar las coordenadas sobre ellos. Las 10 coordenadas sobre el eje 1 se obtienen mediante la multiplicación matricial $\mathbf{X}\mathbf{u}_1$ y sobre el eje 2 mediante $\mathbf{X}\mathbf{u}_2$. En el Gráfico 1-3 se muestra el primer plano factorial (único) del ACP no normado aplicado a la tabla de las 10 muestras por las 2 variables aroma y cuerpo. En este caso, de interés únicamente ilustrativo, la nube de individuos está en \mathbf{R}^2 y el resultado del ACP es la rotación de los ejes. En el gráfico ahora la nube aparece más alargada horizontalmente, ya que el eje 1 se ha colocado en esa dirección. Se muestran los antiguos ejes que representan a las variables aroma y cuerpo.

Gráfico 1-3: Primer plano factorial del ACP aroma-cuerpo



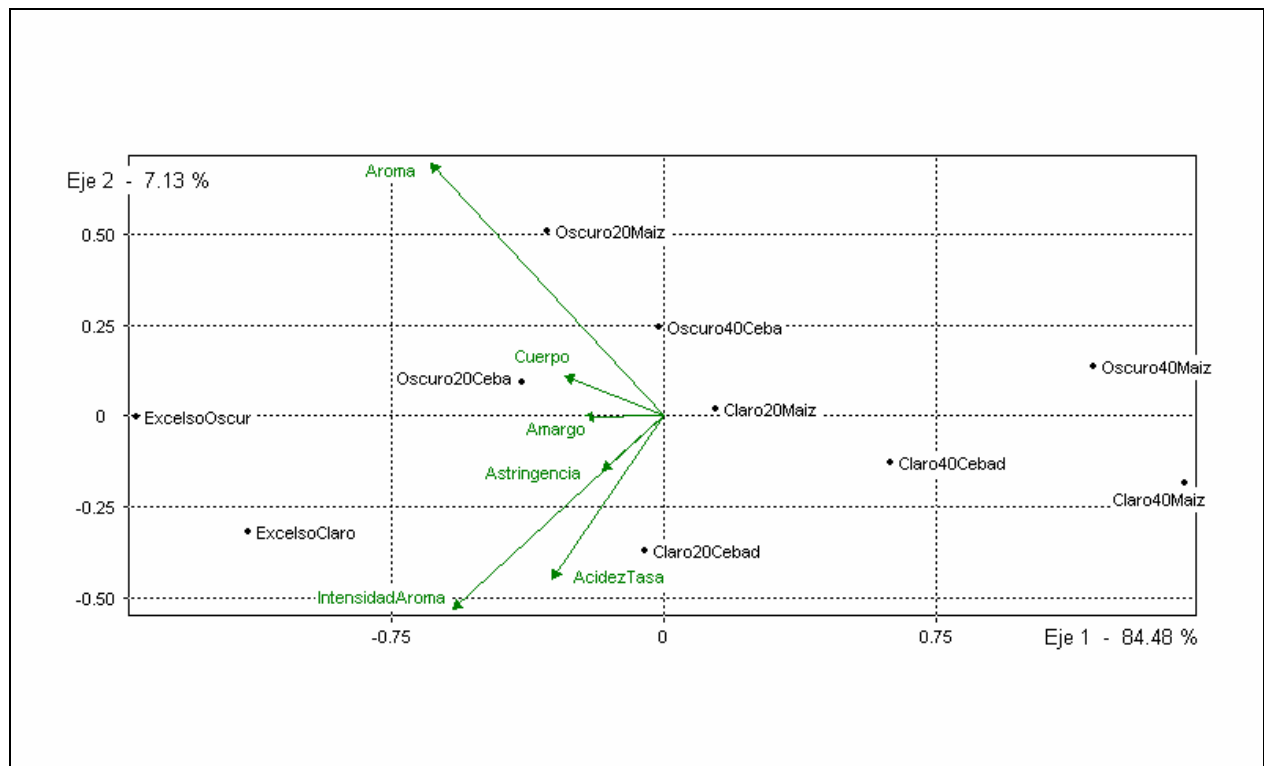
La aplicación del ACP no normado a la Tabla 1-1 completa, implica la representación abstracta de la nube de individuos en \mathbf{R}^6 , la matriz de varianzas y covarianzas está en la Tabla 1-5. En este ACP no normado influyen sobre todo las calificaciones de aroma e intensidad. En la Tabla 1-6 se observa que el primer valor propio es 0.7553 que corresponde al 84.5% de la inercia total. El primer plano acumula el 91.6% de la inercia, entonces la lectura geométrica del primer plano es una muy buena aproximación de la representación en \mathbf{R}^6 , esta aparece en el Gráfico 1-4, en donde además se han proyectado los antiguos ejes de las variables. Rápidamente se puede observar que las marcas excelso son las que obtienen las mejores calificaciones y las contaminadas con 40% de maíz, las peores.

Tabla 1-5: Matriz de covarianzas para la Tabla 1-1

| | Inte | Arom | Cuer | Acid | Amar | Astr |
|------|------|------|------|------|------|------|
| Inte | 0.28 | | | | | |
| Arom | 0.26 | 0.34 | | | | |
| Cuer | 0.12 | 0.13 | 0.07 | | | |
| Acid | 0.14 | 0.13 | 0.06 | 0.10 | | |
| Amar | 0.09 | 0.10 | 0.04 | 0.05 | 0.06 | |
| Astr | 0.08 | 0.08 | 0.03 | 0.04 | 0.03 | 0.04 |

Tabla 1-6: Histograma de valores propios

| NUMERO | VALEUR PROPRE | POURCENT. | POURCENT. CUMULE | |
|--------|---------------|-----------|------------------|-------|
| 1 | 0.7553 | 84.48 | 84.48 | ***** |
| 2 | 0.0637 | 7.13 | 91.61 | ***** |
| 3 | 0.0290 | 3.24 | 94.85 | **** |
| 4 | 0.0221 | 2.47 | 97.33 | *** |
| 5 | 0.0168 | 1.88 | 99.21 | ** |
| 6 | 0.0071 | 0.79 | 100.00 | * |

Gráfico 1-4: Primer plano factorial del ACP no normado de la Tabla 1-1

1.2.3. La nube de variables

La representación geométrica de las variables es menos familiar pero igualmente importante. Conviene pensarla no como puntos sino como flechas que terminan en los puntos correspondientes a las coordenadas de las variables. Lo que interesa observar en las variables es su relación y la analogía geométrica en su representación es el ángulo entre las variables. En el ejemplo del café las variables están representadas en un espacio de 10 dimensiones (\mathbf{R}^{10}), es decir cada individuo es un eje para las variables. Para cada variable sus coordenadas son las respectivas columnas de la Tabla 1-1. Al hacer participar en el cálculo de la distancia el peso de los individuos se obtiene significado geométrico para las diferentes estadísticas de las variables: media, varianza, desviación estándar, correlación. La definición de una distancia particular se introduce al definir un producto interno, a través de una matriz que se le da el nombre de matriz de métrica, \mathbf{M} . El producto interno canónico tiene como matriz de métrica a la identidad, \mathbf{I} , a la que a veces se le agrega un subíndice para indicar su tamaño, por ejemplo \mathbf{I}_n .

La métrica a utilizar para la nube de variables es: $\mathbf{D} = \frac{1}{n} \mathbf{I}_n = \begin{bmatrix} \frac{1}{n} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{n} \end{bmatrix}$

Con lo cual el producto interno entre dos vectores variables \mathbf{r}_j y \mathbf{r}_k , cada uno con n coordenadas se define:

$$\langle \mathbf{r}_j, \mathbf{r}_k \rangle = \mathbf{r}_j' \mathbf{D} \mathbf{r}_k = \frac{1}{n} \sum_{i=1}^n r_{ij} r_{ik}$$

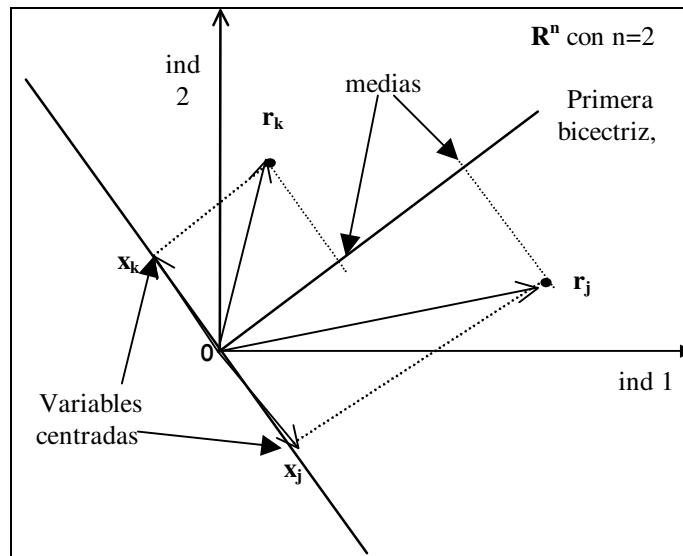
Sea $\mathbf{1}_n$ un vector que tiene todas las n coordenadas iguales a 1, este vector es la primera bisectriz del conjunto de ejes, en el plano es una línea ubicada a 45° de los dos ejes. La media aritmética de una variable se puede expresar:

$$\bar{r}_j = \frac{1}{n} \sum_{i=1}^n r_{ij} = \langle \mathbf{r}_j, \mathbf{1}_n \rangle$$

Entonces la media de una variable está representada como un vector que es la proyección de la variable sobre el vector de unos y tiene entonces n componentes iguales todos al valor de la media de la variable (Gráfico 1-5).

La operación de centrado de un vector variable: $\mathbf{x}_j = \mathbf{r}_j - \bar{r}_j \mathbf{1}_n$, es la proyección del vector sobre un subespacio ortogonal al vector de unos (Gráfico 1-5). Entonces la representación de la nube de variables permanece en el origen luego del centrado, el cual conlleva la pérdida de una dimensión y no el traslado de los ejes.

Gráfico 1-5: Significado de la media y del centrado en la nube de variables



La covarianza es el producto interno entre dos vectores variable centrados:

$$\text{Cov}(r_j, r_k) = \frac{1}{n} \sum_{i=1}^n x_{ij} x_{ik} = \langle \mathbf{x}_j, \mathbf{x}_k \rangle$$

La varianza es la norma al cuadrado del vector variable centrado:

$$\text{Var}(r_j) = \frac{1}{n} \sum_{i=1}^n x_{ij}^2 = \langle \mathbf{x}_j, \mathbf{x}_j \rangle = \|\mathbf{x}_j\|^2$$

La desviación estándar es la norma del vector variable centrado:

$$\sigma_j = \sqrt{\text{Var}(r_j)} = \|\mathbf{x}_j\|$$

La correlación es el coseno del ángulo entre los dos vectores variables centrados:

$$\rho(r_j, r_k) = \frac{Cov(r_j, r_k)}{\sigma_j \sigma_k} = \frac{\langle \mathbf{x}_j, \mathbf{x}_k \rangle}{\|\mathbf{x}_j\| \cdot \|\mathbf{x}_k\|} = \cos(\mathbf{x}_j, \mathbf{x}_k)$$

La nube de variables se puede entender entonces como un conjunto de flechas que parten del origen y son de una longitud igual a la desviación estándar de la variable. El ángulo entre cada par de flechas representa la correlación entre esas dos variables. Ángulos pequeños significan una alta correlación positiva, ángulos cercanos a 90° ausencia de correlación y ángulos cercanos a 180° una alta correlación negativa.

Una aproximación de esa representación es la que se obtiene en el primer plano factorial de las variables. Para el ejemplo de las muestras de café el primer plano factorial de las variables se muestra en el Gráfico 1-6, donde se observan las variables aroma e intensidad de aroma como las de más desviación estándar y por lo tanto las de mayor influencia en el análisis. También se observa que todas las variables apuntan al lado negativo del primer eje, indicando de una parte una fuerte relación entre todas las variables y de otra indicando que el lado negativo del primer eje indica mayores calificaciones en todos los aspectos de la degustación. En la Tabla 1-7 aparecen los números que permiten construir el Gráfico 1-6 (coordenadas), las correlaciones entre los factores y las variables (aroma e intensidad de aroma son las más correlacionadas con el primer eje) y los vectores propios (lectura vertical de los antiguos ejes unitarios) que dan las direcciones de los nuevos ejes sobre el sistema de representación original de la nube de individuos. Los antiguos ejes unitarios leídos de forma horizontal permiten ubicar los ejes de las variables en el plano de los individuos (Gráfico 1-4), por ejemplo *IntensidadAroma* se ubica en el primer plano con las coordenadas -0.58 y -0.53 (parte inferior izquierda)

Gráfico 1-6: Primer plano factorial de las variables del ACP no normado de las muestras de café

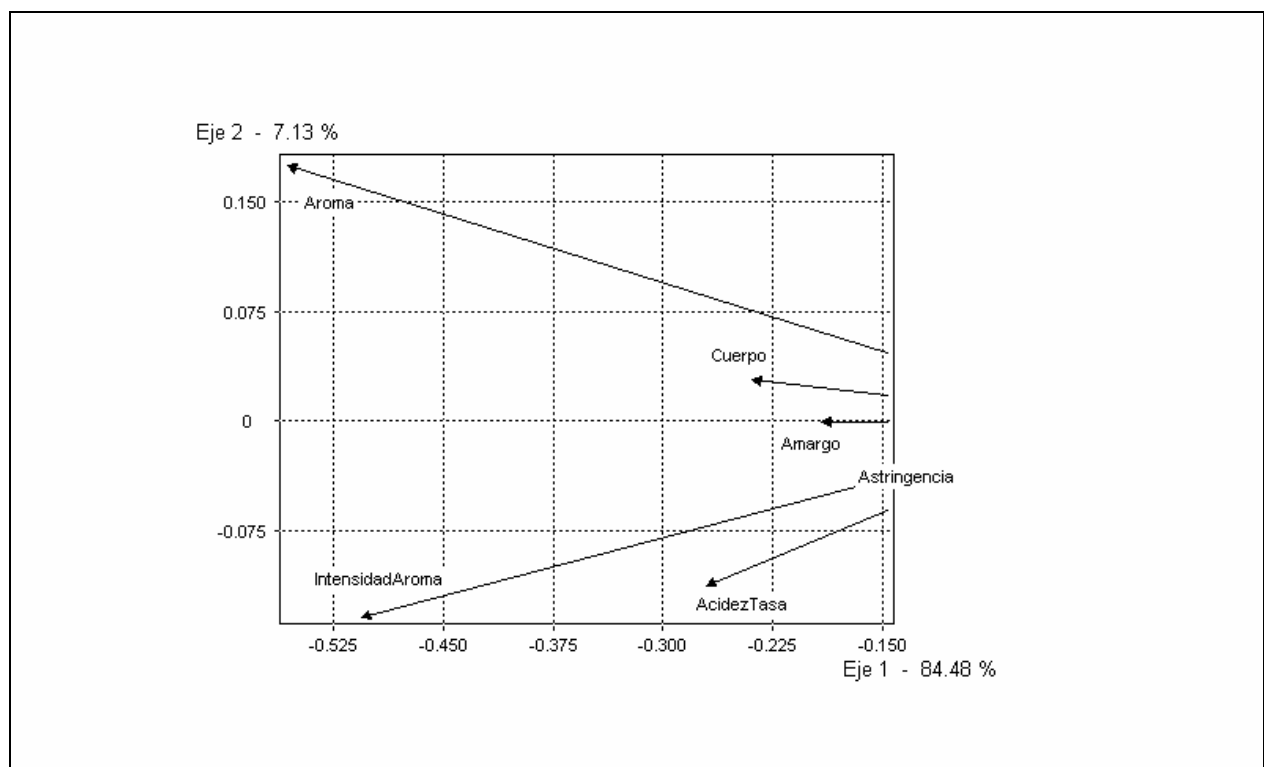


Tabla 1-7: Coordenadas, correlaciones con los ejes y antiguos ejes unitarios para las variables

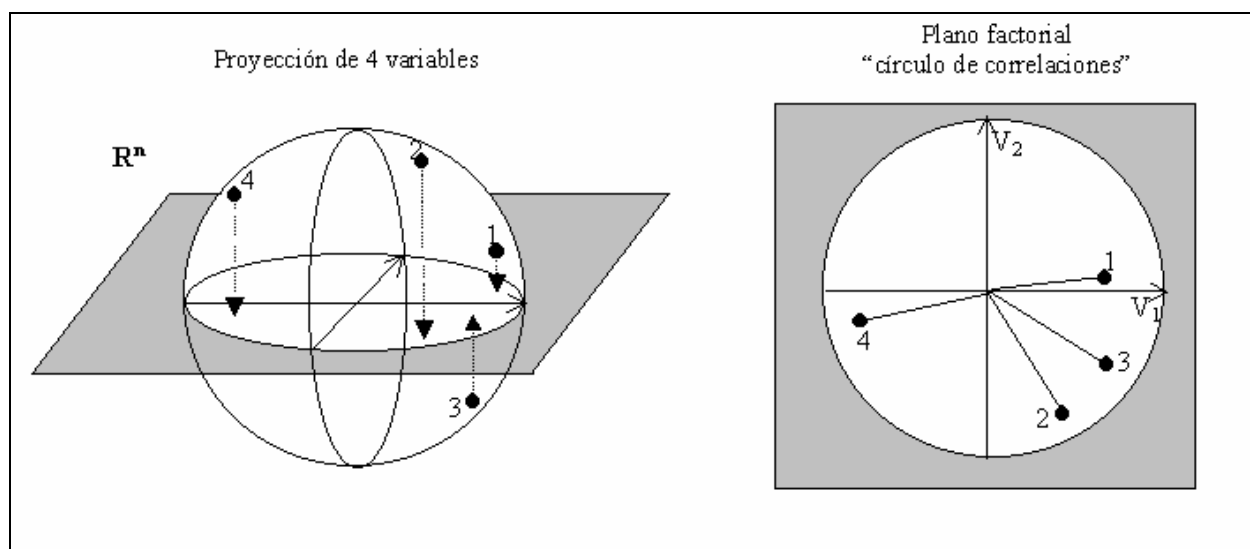
| VARIABLES | COORDONNEES | | | | | CORRELATIONS VARIABLE-FACTEUR | | | | | ANCIENS AXES UNITAIRES | | | | |
|------------------------|-------------|-------|-------|-------|-------|-------------------------------|-------|-------|-------|-------|------------------------|-------|-------|-------|-------|
| IDEN - LIBELLE COURT | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| Inte - IntensidadAroma | -0.51 | -0.13 | 0.06 | -0.02 | 0.01 | -0.96 | -0.25 | 0.11 | -0.04 | 0.02 | -0.58 | -0.53 | 0.33 | -0.14 | 0.08 |
| Arom - Aroma | -0.56 | 0.18 | 0.00 | 0.04 | -0.03 | -0.95 | 0.30 | 0.00 | 0.06 | -0.05 | -0.64 | 0.69 | 0.00 | 0.24 | -0.21 |
| Cuer - Cuerpo | -0.24 | 0.03 | 0.04 | -0.03 | 0.09 | -0.89 | 0.11 | 0.15 | -0.13 | 0.35 | -0.27 | 0.11 | 0.24 | -0.23 | 0.71 |
| Acid - AcidezTasa | -0.27 | -0.11 | -0.07 | 0.09 | 0.00 | -0.85 | -0.36 | -0.22 | 0.29 | -0.01 | -0.31 | -0.45 | -0.41 | 0.61 | -0.02 |
| Amar - Amargo | -0.19 | 0.00 | -0.13 | -0.09 | 0.01 | -0.77 | 0.00 | -0.53 | -0.35 | 0.03 | -0.22 | 0.00 | -0.77 | -0.59 | 0.05 |
| Astr - Astringencia | -0.15 | -0.04 | 0.04 | -0.06 | -0.09 | -0.76 | -0.19 | 0.22 | -0.31 | -0.44 | -0.17 | -0.15 | 0.26 | -0.40 | -0.66 |

1.2.4. ACP normado

En el ACP normado las coordenadas de los individuos son los valores estandarizados de cada una de las variables. Una variable estandarizada pierde las unidades de medida con lo cual se elimina el efecto de la escala de medida. El valor de una variable estandarizada corresponde para un individuo a su distancia al promedio expresada en el número de desviaciones estándar. Como se sabe la desviación estándar de una variable estandarizada es igual a uno y los valores de las covarianzas entre variables estandarizadas son iguales a sus coeficientes de correlación. Con la estandarización todas las variables contribuyen igual a la inercia (con 1), de modo que la inercia total es igual al número de variables.

La nube de individuos cambia con respecto al no normado pero su interpretación sigue siendo la misma. La nube de variables tiene en el caso normado una representación más sencilla. Al tener todas las variables la misma longitud igual a 1 (su desviación estándar), la nube de variables corresponde a puntos sobre un cascarón hipersférico de radio uno (Gráfico 1-7). El primer plano factorial corresponde al plano que corta a la hipersfera de tal manera que se logra la mejor proyección de las variables. Una variable que quede exactamente sobre el plano factorial tendrá longitud 1 en su proyección y su flecha terminará en un círculo de radio uno, que se denomina círculo de correlaciones.

Gráfico 1-7: Ejemplo de representación de 4 variables y su proyección



(Fuente: Lebart et al., 1995)

Desde el punto de vista de las variables el primer plano factorial del ACP normado es el mejor resumen gráfico de la matriz de correlaciones (Gráfico 1-7). Las variables más correlacionadas entre sí mostrarán un

ángulo pequeño y las menores un ángulo cercano a los noventa grados. Las variables correlacionadas negativamente tendrán direcciones opuestas en el plano. Las variables bien representadas en el plano se aproximan al círculo de radio 1. Para una variable bien representada en el plano un ángulo pequeño con el eje indica una buena representación con el el. La coordenada de una variable sobre un eje factorial corresponde a la representación con ese eje (el eje representa una nueva variable y las coordenadas de los individuos son los valores que toma).

El ACP normado es el que más se utiliza porque, en la práctica, las variables suelen tener diferente escala aún en los casos en que las unidades de medida sean las mismas. Por ejemplo si se está midiendo lo que gasta un hogar en comida, se puede llevar todo a pesos, pero las escalas siguen siendo distintas por el orden de las cantidades consumidas y de los precios de los productos. (por ejemplo carne vs. pan).

En el caso de la calificación de las muestras de café dadas por los degustadores se observan que suelen dan mayor calificación a unos aspectos que a otros, con el ACP normado se hace jugar un papel similar a todos. La Tabla 1-8 es la matriz de correlaciones, las cuales son todas positivas y elevadas. El Gráfico 1-8 es el círculo de correlaciones, esta imagen es típica cuando todas las correlaciones son positivas, el primer eje recoge la mayoría de la inercia (76.7% Tabla 1-9), constituyéndose en un “factor tamaño”, suficiente para resumir la información de las 6 variables. Al lado izquierdo están entonces las muestras de café mejor calificadas (Gráfico 1-9). El eje dos, de importancia muy inferior al primero separa las muestras de mayor astringencia y menor amargo (arriba) con mayor amargo y menor astringencia (abajo). La Tabla 1-10 contiene la información para construir el Gráfico 1-8, las columnas debajo de antiguos ejes unitarios corresponden a los cinco vectores propios asociados a los cinco mayores valores propios, que aparecen en la Tabla 1-9. Con los coeficientes de la primera columna se construye la fórmula de la primera componente principal, variable sintética que resume la información de las seis variables y que se puede denominar “índice de calidad apreciada”. La Tabla 1-11 contiene la información para la construcción del Gráfico 1-9 e índices complementarios para su lectura. La contribución indica los individuos que más aportan a la construcción de cada uno de los ejes. El coseno cuadrado es un indicador de la calidad de la proyección de un individuo sobre el eje. La suma del coseno cuadrado de los ejes 1 y 2 corresponde a la calidad de la representación sobre le primer plano factorial.

Tabla 1-8: Matriz de correlaciones

| | Inte | Arom | Cuer | Acid | Amar | Astr |
|------|------|------|------|------|------|------|
| Inte | 1.00 | | | | | |
| Arom | 0.83 | 1.00 | | | | |
| Cuer | 0.84 | 0.86 | 1.00 | | | |
| Acid | 0.87 | 0.72 | 0.67 | 1.00 | | |
| Amar | 0.70 | 0.71 | 0.66 | 0.67 | 1.00 | |
| Astr | 0.78 | 0.66 | 0.62 | 0.61 | 0.56 | 1.00 |

Tabla 1-9: Histograma de los valores propios del ACP normado “café”

| NUMERO | VALEUR PROPRE | POURCENT. PROPRE | POURCENT. CUMULE | |
|--------|------------------|---------------------|---------------------|-------|
| 1 | 4.6015 | 76.69 | 76.69 | ***** |
| 2 | 0.4694 | 7.82 | 84.51 | ***** |
| 3 | 0.3845 | 6.41 | 90.92 | ***** |
| 4 | 0.3446 | 5.74 | 96.67 | ***** |
| 5 | 0.1449 | 2.41 | 99.08 | *** |
| 6 | 0.0551 | 0.92 | 100.00 | * |

Tabla 1-10: Coordenadas y antiguos ejes unitarios del ACP normado “café”

| VARIABLES | COORDONNEES | | | | | CORRELATIONS VARIABLE-FACTEUR | | | | | ANCIENS AXES UNITAIRES | | | | |
|------------------------|-------------|-------|-------|-------|-------|-------------------------------|-------|-------|-------|-------|------------------------|-------|-------|-------|-------|
| IDEN - LIBELLE COURT | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| Inte - IntensidadAroma | -0.96 | 0.10 | 0.01 | 0.15 | -0.10 | -0.96 | 0.10 | 0.01 | 0.15 | -0.10 | -0.45 | 0.15 | 0.01 | 0.26 | -0.26 |
| Arom - Aroma | -0.91 | -0.10 | -0.25 | -0.05 | 0.30 | -0.91 | -0.10 | -0.25 | -0.05 | 0.30 | -0.43 | -0.14 | -0.40 | -0.08 | 0.79 |
| Cuer - Cuerpo | -0.89 | -0.10 | -0.39 | 0.00 | -0.20 | -0.89 | -0.10 | -0.39 | 0.00 | -0.20 | -0.41 | -0.14 | -0.63 | 0.00 | -0.53 |
| Acid - AcidezTasa | -0.87 | -0.07 | 0.27 | 0.40 | 0.04 | -0.87 | -0.07 | 0.27 | 0.40 | 0.04 | -0.40 | -0.10 | 0.44 | 0.68 | 0.09 |
| Amar - Amargo | -0.82 | -0.37 | 0.29 | -0.34 | -0.05 | -0.82 | -0.37 | 0.29 | -0.34 | -0.05 | -0.38 | -0.53 | 0.47 | -0.58 | -0.13 |
| Astr - Astringencia | -0.80 | 0.55 | 0.11 | -0.21 | 0.01 | -0.80 | 0.55 | 0.11 | -0.21 | 0.01 | -0.37 | 0.80 | 0.18 | -0.36 | 0.02 |

Tabla 1-11: Coordenadas, contribuciones y cosenos cuadrados de las muestras de café (ACP normado)

| INDIVIDUS | | | COORDONNEES | | | | | CONTRIBUTIONS | | | | | COSINUS CARRES | | | | |
|----------------|-------|-------|-------------|-------|-------|-------|-------|---------------|------|------|------|------|----------------|------|------|------|------|
| IDENTIFICATEUR | P.REL | DISTO | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| ExcelsoClaro | 10.00 | 10.41 | -2.99 | 1.02 | 0.61 | 0.04 | 0.03 | 19.5 | 22.0 | 9.6 | 0.0 | 0.1 | 0.86 | 0.10 | 0.04 | 0.00 | 0.00 |
| Claro40Maiz | 10.00 | 10.95 | 3.21 | 0.04 | 0.78 | 0.07 | -0.09 | 22.3 | 0.0 | 15.8 | 0.1 | 0.6 | 0.94 | 0.00 | 0.06 | 0.00 | 0.00 |
| Claro40Cebad | 10.00 | 2.71 | 1.39 | -0.56 | 0.66 | 0.18 | -0.03 | 4.2 | 6.6 | 11.4 | 1.0 | 0.1 | 0.71 | 0.11 | 0.16 | 0.01 | 0.00 |
| Claro20Maiz | 10.00 | 4.03 | 0.98 | 0.22 | -1.03 | 1.39 | 0.15 | 2.1 | 1.0 | 27.4 | 56.3 | 1.6 | 0.24 | 0.01 | 0.26 | 0.48 | 0.01 |
| Claro20Cebad | 10.00 | 1.10 | -0.35 | -0.50 | 0.46 | 0.41 | -0.58 | 0.3 | 5.4 | 5.6 | 4.9 | 22.9 | 0.11 | 0.23 | 0.20 | 0.15 | 0.30 |
| ExcelsoOscur | 10.00 | 12.71 | -3.54 | -0.09 | 0.27 | 0.10 | 0.29 | 27.2 | 0.2 | 1.8 | 0.3 | 5.8 | 0.99 | 0.00 | 0.01 | 0.00 | 0.01 |
| Oscuro40Maiz | 10.00 | 10.93 | 3.02 | 1.11 | -0.15 | -0.65 | 0.29 | 19.8 | 26.4 | 0.6 | 12.1 | 5.7 | 0.83 | 0.11 | 0.00 | 0.04 | 0.01 |
| Oscuro40Ceba | 10.00 | 2.00 | 0.14 | -0.85 | -0.71 | -0.72 | -0.33 | 0.0 | 15.3 | 13.1 | 15.1 | 7.4 | 0.01 | 0.36 | 0.25 | 0.26 | 0.05 |
| Oscuro20Maiz | 10.00 | 1.83 | -0.50 | -0.91 | -0.16 | -0.36 | 0.76 | 0.5 | 17.5 | 0.7 | 3.7 | 39.7 | 0.14 | 0.45 | 0.01 | 0.07 | 0.31 |
| Oscuro20Ceba | 10.00 | 3.32 | -1.35 | 0.52 | -0.73 | -0.47 | -0.49 | 4.0 | 5.7 | 13.9 | 6.4 | 16.2 | 0.55 | 0.08 | 0.16 | 0.07 | 0.07 |

Gráfico 1-8: Círculo de correlación para el ejemplo café

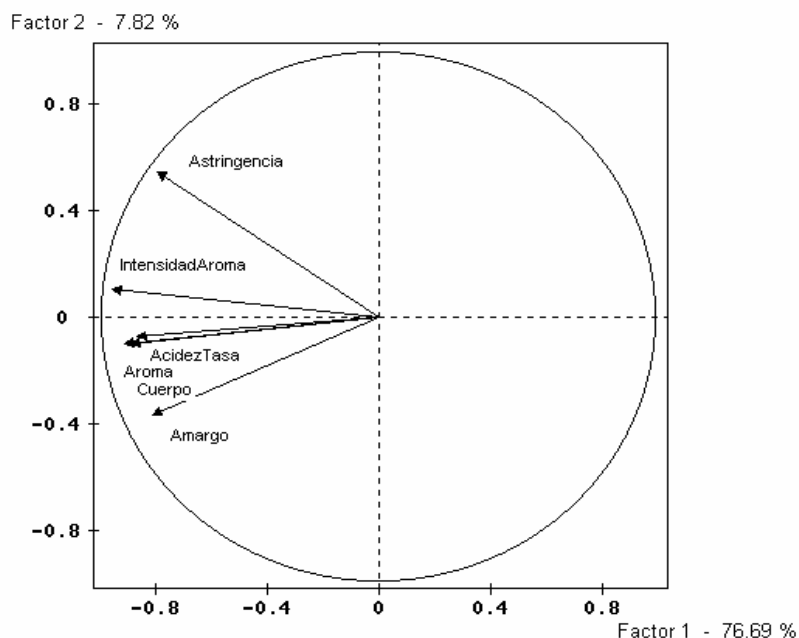
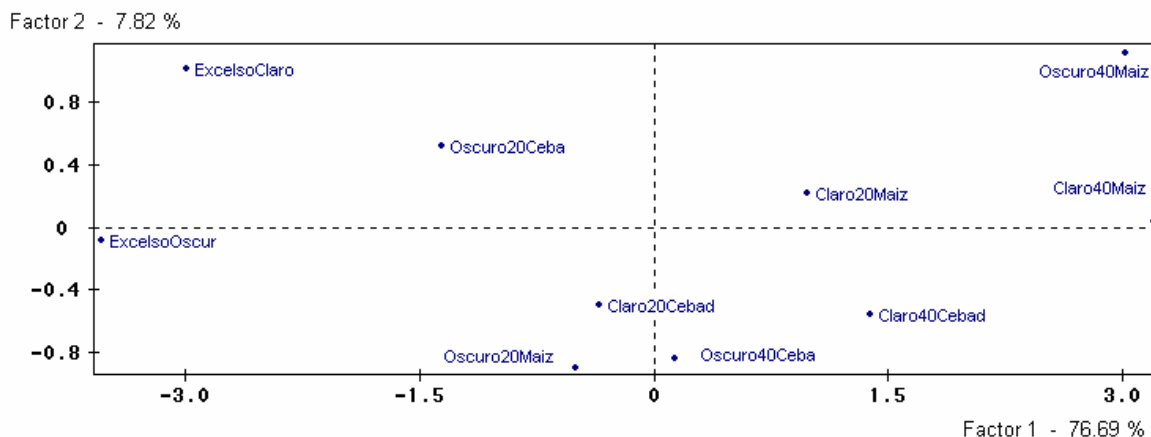


Gráfico 1-9: Primer plano factorial del ACP normado del ejemplo café

1.2.5. Elementos suplementarios

Es posible proyectar elementos suplementarios o ilustrativos (individuos, variables continuas y variables nominales) sobre los planos construidos en el ACP. Los elementos suplementarios permiten explorar asociaciones con los elementos activos o enriquecer los análisis. Los elementos que participan en el análisis se denominan activos, en el caso del ACP son variables continuas activas e individuos activos.

Individuos suplementarios

Un individuo que tiene los valores para todas las variables activas pero que no participó en el ACP se puede proyectar sobre los ejes obtenidos de la misma forma que los individuos activos. Mediante este procedimiento se puede posicionar un nuevo individuo con respecto a todos los individuos activos para responder a objetivos preestablecidos, por ejemplo explorar su posible discriminación entre grupos.

En el ejemplo del café se han degustado también dos marcas comerciales y su posición como ilustrativas indica su calidad apreciada (Gráfico 1-10). El café comercial 2 obtiene una nota más alta que el comercial 1.

Variables nominales ilustrativas

Como una variable nominal representa una partición (división en clases) de los individuos lo que se proyecta son los centros de gravedad de cada subconjunto asociado a una modalidad. La variable nominal agregado (nada, maíz, cebada) se proyecta ubicando los individuos promedio de cada uno de los tres grupos, valores que se muestran en la Tabla 1-12 y que se representan en el Gráfico 1-10. Se observa que las muestras sin agregado de cereal son las de mejor calificación seguida de las de agregado de avena, con calificación intermedia y finalmente las de maíz que son las de calificación inferior. En la Tabla 1-13 aparecen las coordenadas y los índices que ayudan a la interpretación de las modalidades proyectadas.

Adicionalmente se muestra la frecuencia, el peso y la distancia al cuadrado al centro de gravedad de cada modalidad. El valor test es un indicador para decidir sobre la lectura de una modalidad proyectada respecto al origen de la representación. Se construye en base a una prueba de hipótesis (H_0 : la coordenada del centro de gravedad de la modalidad se diferencia de cero solamente por fluctuaciones aleatorias). En general los valores p de las pruebas de hipótesis se recodifican a cuantiles de una distribución normal estándar, es decir que valores superiores a 2 o a -2 indican que vale la pena leer la modalidad con respecto a ese eje.

Las modalidades nada (valor test = -2.3) y maíz (valor test = 1.9) se pueden considerar de coordenadas diferentes a cero en el primer eje, pero la modalidad avena no (valor test = 0).

Variables continuas

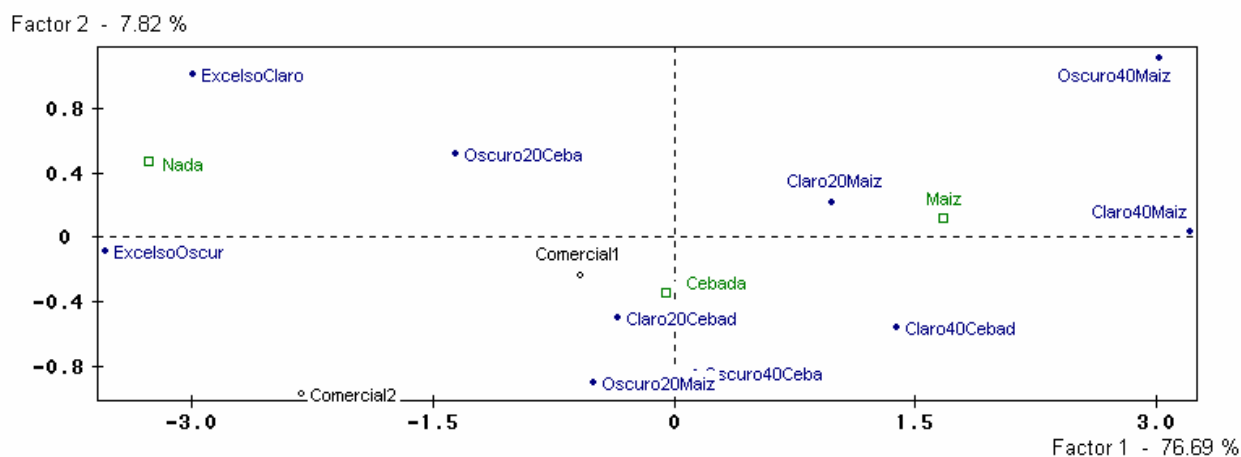
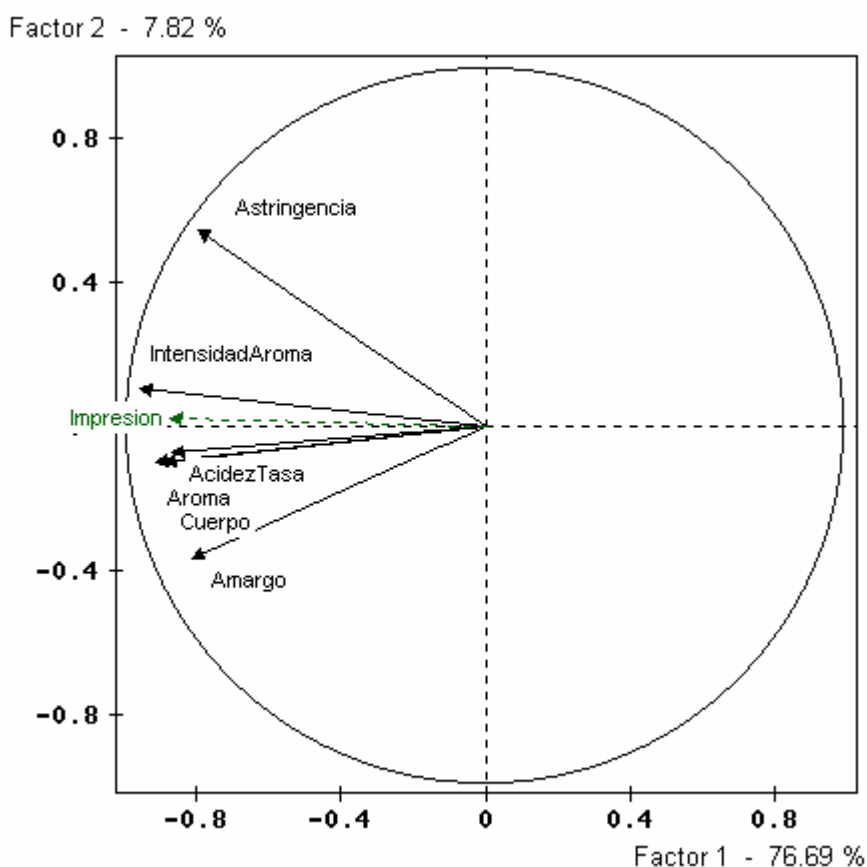
En el ACP normado la proyección de una variable continua ilustrativa equivale a su correlación con el eje, lo que da la clave para su interpretación. La variable impresión global proyectada como ilustrativa se muestra en el Gráfico 1-11, mostrando que esta calificación se correlaciona bastante bien con el primer eje.

Tabla 1-12: Tabla de datos mostrando los promedios según las modalidades de agregado

| IDEN | Agredado | Intensidad Aroma | Aroma | Cuerpo | AcidezTasa | Amargo | Astringenci a | Impresion |
|-----------------------|----------|---------------------|-------------|-------------|-------------|-------------|------------------|-----------|
| ExcelsoClaro | Nada | 7.72 | 7.00 | 6.84 | 5.02 | 5.04 | 5.36 | 7.46 |
| ExcelsoOscur | Nada | 7.66 | 7.42 | 6.98 | 5.12 | 5.18 | 5.22 | 7.40 |
| PromedioNada | | 7.69 | 7.21 | 6.91 | 5.07 | 5.11 | 5.29 | |
| Claro40Maiz | Maiz | 6.02 | 5.42 | 6.22 | 4.34 | 4.60 | 4.78 | 6.24 |
| Claro20Maiz | Maiz | 6.82 | 6.44 | 6.70 | 4.62 | 4.38 | 4.80 | 6.04 |
| Oscuro40Maiz | Maiz | 6.18 | 5.82 | 6.26 | 4.00 | 4.46 | 4.96 | 5.90 |
| Oscuro20Maiz | Maiz | 6.66 | 7.06 | 6.70 | 4.64 | 5.00 | 4.90 | 6.90 |
| PromedioMaiz | | 6.42 | 6.19 | 6.47 | 4.40 | 4.61 | 4.86 | |
| Claro40Cebad | Cebada | 6.48 | 5.98 | 6.44 | 4.58 | 4.82 | 4.80 | 6.12 |
| Claro20Cebad | Cebada | 7.08 | 6.20 | 6.72 | 4.78 | 4.94 | 4.90 | 6.22 |
| Oscuro40Ceba | Cebada | 6.84 | 6.56 | 6.82 | 4.30 | 4.96 | 4.84 | 6.94 |
| Oscuro20Ceba | Cebada | 7.00 | 6.70 | 7.04 | 4.60 | 4.88 | 5.18 | 7.16 |
| PromedioCebada | | 6.85 | 6.36 | 6.76 | 4.57 | 4.90 | 4.93 | |
| Comercial1 | | 6.48 | 5.46 | 7.06 | 4.60 | 5.26 | 5.16 | 5.68 |
| Comercial2 | | 6.66 | 6.26 | 7.36 | 4.44 | 5.64 | 5.24 | 6.00 |

Tabla 1-13: Coordenadas y valores test de las modalidades

| MODALITES | | | VALEURS-TEST | | | | | COORDONNEES | | | | | |
|----------------|------|--------|--------------|------|------|------|------|-------------|-------|-------|-------|-------|--------|
| IDEN - LIBELLE | EFF. | P. ABS | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | DISTO. |
| 2 . Agredado | | | | | | | | | | | | | |
| Maiz - Maiz | 4 | 4.00 | 1.9 | 0.4 | -0.6 | 0.5 | 1.8 | 1.68 | 0.12 | -0.14 | 0.12 | 0.28 | 2.94 |
| Ceba - Cebada | 4 | 4.00 | 0.0 | -1.2 | -0.3 | -0.6 | -2.3 | -0.04 | -0.35 | -0.08 | -0.15 | -0.36 | 0.28 |
| Nada - Nada | 2 | 2.00 | -2.3 | 1.0 | 1.1 | 0.2 | 0.6 | -3.27 | 0.46 | 0.44 | 0.07 | 0.16 | 11.13 |

Gráfico 1-10: Primer plano factorial con individuos y modalidades ilustrativas**Gráfico 1-11: Impresión proyectada como variable continua ilustrativa**

1.3. Un ejemplo de aplicación: perfil productivo de los investigadores

1.3.1. Presentación

En un análisis de la Convocatoria de grupos y centros de investigación realizado por Colciencias en 1998 (Charum et al.) se identificaron 2148 investigadores, para los que se tiene, entre otras informaciones, cantidades de resultados producidos por cada uno de ellos (artículos de revistas, libros, etc). Una primera exploración multivariada de la producción se hizo utilizando análisis en componentes principales buscando la asociación con otras características de los investigadores.

En la Tabla 1-14 aparecen las variables continuas activas, en la primera parte, y las ilustrativas, las tres últimas, que se utilizan en el ACP, junto con sus estadísticas básicas. Por ejemplo para producción de artículos por investigador el mínimo es 0 y el máximo 67, el promedio es 2.90 y la desviación estándar 5.10. Las variables activas son las cantidades de los diferentes tipos de producto que reportaron los investigadores durante el periodo de análisis. El ACP que se utiliza aquí es el normado es decir que las variables originales se estandarizan.

1.3.2. Análisis de tablas y gráficos

La Tabla 1-15 es la matriz de correlaciones, debajo aparecen los valores tests asociados a las hipótesis nulas de no-correlación. La decisión para saber cuántos ejes es conveniente analizar está soportada en el histograma de valores propios (Tabla 1-16), en este caso se decide utilizar los dos primeros ejes para la síntesis del perfil productivo de los investigadores (40% de la inercia). Sobre el círculo de correlaciones se pueden proyectar variables continuas ilustrativas, para establecer para cada una de ellas su asociación con las variables activas. La Tabla 1-17 apoya la lectura del gráfico de correlaciones (Gráfico 1-12), incluyendo variables activas e ilustrativas. La correlación entre el número total de productos y el primer eje (-0.91) indica que ambas variables, el total y el primer componente principal (la fórmula para obtener las proyecciones sobre un eje es la expresión de esta nueva variable), están midiendo más o menos lo mismo, y sirven como un índice univariado de toda la producción.

Tabla 1-14: Estadísticas básicas de las variables continuas del ACP

| NUM . IDEN - ETIQUETA | FRECUENCIA | PESOS | MEDIA | DESV. EST. | MINIMO | MAXIMO |
|----------------------------|------------|---------|-------|------------|--------|--------|
| 19 . Arti - Artículo | 2148 | 2148.00 | 2.90 | 5.10 | 0.00 | 67.00 |
| 20 . Libr - LibroAutor | 2148 | 2148.00 | 0.28 | 0.82 | 0.00 | 11.00 |
| 21 . CapL - CapLibro | 2148 | 2148.00 | 0.23 | 0.82 | 0.00 | 10.00 |
| 22 . CapC - CapCompendio | 2148 | 2148.00 | 0.09 | 0.72 | 0.00 | 18.00 |
| 23 . CapM - CapMemoria | 2148 | 2148.00 | 0.41 | 1.63 | 0.00 | 23.00 |
| 24 . Tesi - Tesis | 2148 | 2148.00 | 0.27 | 0.96 | 0.00 | 15.00 |
| 25 . Otro - OtroProducto | 2148 | 2148.00 | 0.14 | 0.59 | 0.00 | 8.00 |
| 1 . Edad - Edad | 2082 | 2082.00 | 43.26 | 9.43 | 22.00 | 78.00 |
| 9 . NoId - NoIdiomasEx | 2148 | 2148.00 | 0.79 | 0.84 | 0.00 | 6.00 |
| 26 . Tota - TotalProductos | 2148 | 2148.00 | 4.32 | 6.47 | 0.00 | 70.00 |

Tabla 1-15: Matriz de correlaciones y valores test para la producción de los investigadores

| | Arti | Libr | CapL | CapC | CapM | Tesi | Otro |
|------|------|-------|------|-------|------|------|------|
| Arti | 1.00 | | | | | | |
| Libr | 0.06 | 1.00 | | | | | |
| CapL | 0.25 | 0.22 | 1.00 | | | | |
| CapC | 0.11 | 0.04 | 0.10 | 1.00 | | | |
| CapM | 0.16 | 0.02 | 0.11 | -0.01 | 1.00 | | |
| Tesi | 0.19 | -0.02 | 0.06 | 0.01 | 0.12 | 1.00 | |
| Otro | 0.08 | 0.02 | 0.02 | -0.02 | 0.21 | 0.08 | 1.00 |

| MATRIZ DE VALORES-TESTS | | | | | | | |
|-------------------------|-------|-------|-------|-------|-------|-------|-------|
| | Arti | Libr | CapL | CapC | CapM | Tesi | Otro |
| Arti | 99.99 | | | | | | |
| Libr | 2.85 | 99.99 | | | | | |
| CapL | 11.66 | 10.31 | 99.99 | | | | |
| CapC | 4.94 | 2.00 | 4.65 | 99.99 | | | |
| CapM | 7.47 | 1.05 | 4.89 | -0.27 | 99.99 | | |
| Tesi | 8.97 | -0.78 | 2.63 | 0.66 | 5.60 | 99.99 | |
| Otro | 3.82 | 0.79 | 0.94 | -0.80 | 9.65 | 3.53 | 99.99 |

Tabla 1-16: Valores propios del ACP e histograma

| NUMERO | VALOR PROPIO | PORCENT. | PORCENT. ACUMUL. | |
|--------|--------------|----------|------------------|-------|
| 1 | 1.5808 | 22.58 | 22.58 | ***** |
| 2 | 1.1952 | 17.07 | 39.66 | ***** |
| 3 | 1.0188 | 14.55 | 54.21 | ***** |
| 4 | 0.9319 | 13.31 | 67.52 | ***** |
| 5 | 0.8224 | 11.75 | 79.27 | ***** |
| 6 | 0.7662 | 10.95 | 90.22 | ***** |
| 7 | 0.6847 | 9.78 | 100.00 | ***** |

En el Gráfico 1-12 se observa que el primer eje es un factor tamaño, puesto que las coordenadas hacia la izquierda indican mayor producción de los diferentes tipos. En el primer eje tienen mayor ponderación las variables *artículo*, *capítulo de libro* y *capítulo de memoria* (Tabla 1-17). El segundo eje separa: hacia arriba más producción de *libros de autor*, *capítulo de libro* y *capítulo de compendio* y, hacia abajo, *capítulo memoria*, *otro producto* (representa a los productos: diseño industrial, norma, patente, prototipo industrial, software y variedad biológica) y *tesis*.

Se proyectan también variables nominales ilustrativas sobre el primer plano factorial de los individuos, en este caso lo que hace el método es obtener individuos típicos (promedio de todas las variables) de los grupos generados por cada una de las modalidades y proyectarlos como ilustrativos. La Tabla 1-18 apoya la lectura de las modalidades ilustrativas sobre el primer plano factorial de los individuos que se presenta en el Gráfico 1-13. Los valores tests de cada modalidad indican si la coordenada sobre el eje se puede considerar diferente de cero lo cual hace válida la lectura sobre el plano con la relación al respectivo eje. Por ejemplo la modalidad *pregrado* de la variable *máxima escolaridad* tiene un valor test de 5.0, indicando que esta del lado positivo del primer eje, es decir los investigadores de *pregrado* tienen en promedio menor producción. Lo contrario ocurre con el grupo de investigadores con título de doctorado.

Se observa (Gráfico 1-13), que *ciencias básicas* es el área de trabajo de más producción, con proporción mayor que el promedio general en *tesis* y *capítulos de memoria*. Los investigadores del área de *ingeniería* y *tecnología* producen más *otros productos*; *ciencias sociales* y *humanas* aparecen en general con menos producción que el promedio pero con más producción en *libros de autor*, *capítulos de libro* y *compendios*. La producción está relacionada con la escolaridad de los investigadores: los que más producen son los que tienen el nivel de doctorado y los que menos el de nivel de pregrado.

Tabla 1-17: Coordenadas y antiguos ejes unitarios para las variables continuas

| VARIABLES | COORDENADAS | | | | | CORRELACIONES VARIABLE-FACTOR | | | | | ANTIGUOS EJES UNITARIOS | | | | |
|-------------------------|-------------|-------|-------|-------|-------|-------------------------------|-------|-------|-------|-------|-------------------------|-------|-------|-------|-------|
| IDEN - ETIQUETA | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| Arti - Artículo | -0.67 | 0.03 | 0.29 | -0.13 | 0.27 | -0.67 | 0.03 | 0.29 | -0.13 | 0.27 | -0.54 | 0.03 | 0.29 | -0.14 | 0.30 |
| Libr - LibroAutor | -0.34 | 0.52 | -0.54 | -0.13 | -0.44 | -0.34 | 0.52 | -0.54 | -0.13 | -0.44 | -0.27 | 0.48 | -0.53 | -0.14 | -0.49 |
| CapL - CapLibro | -0.60 | 0.44 | -0.13 | -0.14 | 0.27 | -0.60 | 0.44 | -0.13 | -0.14 | 0.27 | -0.48 | 0.40 | -0.13 | -0.15 | 0.30 |
| CapC - CapCompendio | -0.25 | 0.39 | 0.47 | 0.71 | -0.19 | -0.25 | 0.39 | 0.47 | 0.71 | -0.19 | -0.20 | 0.36 | 0.47 | 0.73 | -0.21 |
| CapM - CapMemoria | -0.52 | -0.44 | -0.22 | 0.16 | 0.33 | -0.52 | -0.44 | -0.22 | 0.16 | 0.33 | -0.42 | -0.40 | -0.22 | 0.17 | 0.36 |
| Tesi - Tesis | -0.43 | -0.33 | 0.46 | -0.43 | -0.50 | -0.43 | -0.33 | 0.46 | -0.43 | -0.50 | -0.34 | -0.30 | 0.45 | -0.44 | -0.55 |
| Otro - OtroProducto | -0.36 | -0.52 | -0.38 | 0.41 | -0.30 | -0.36 | -0.52 | -0.38 | 0.41 | -0.30 | -0.29 | -0.48 | -0.37 | 0.42 | -0.33 |
| VARIABLES ILLUSTRATIVES | | | | | | | | | | | | | | | |
| Edad - Edad | -0.12 | 0.11 | 0.01 | -0.05 | -0.02 | -0.12 | 0.11 | 0.01 | -0.05 | -0.02 | | | | | |
| NoId - NoIdiomasEx | -0.13 | 0.04 | 0.05 | 0.03 | 0.00 | -0.13 | 0.04 | 0.05 | 0.03 | 0.00 | | | | | |
| Tota - TotalProductos | -0.91 | -0.02 | 0.18 | -0.05 | 0.15 | -0.91 | -0.02 | 0.18 | -0.05 | 0.15 | | | | | |

Gráfico 1-12: Circulo de correlaciones (primer plano factorial de las variables del ACP)

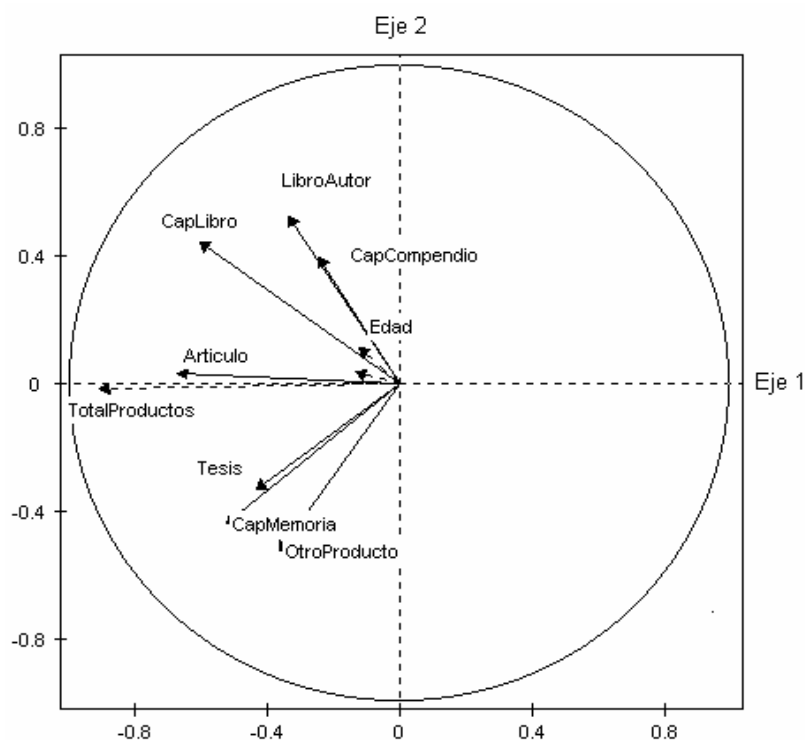


Gráfico 1-13: Primer plano factorial con modalidades ilustrativas

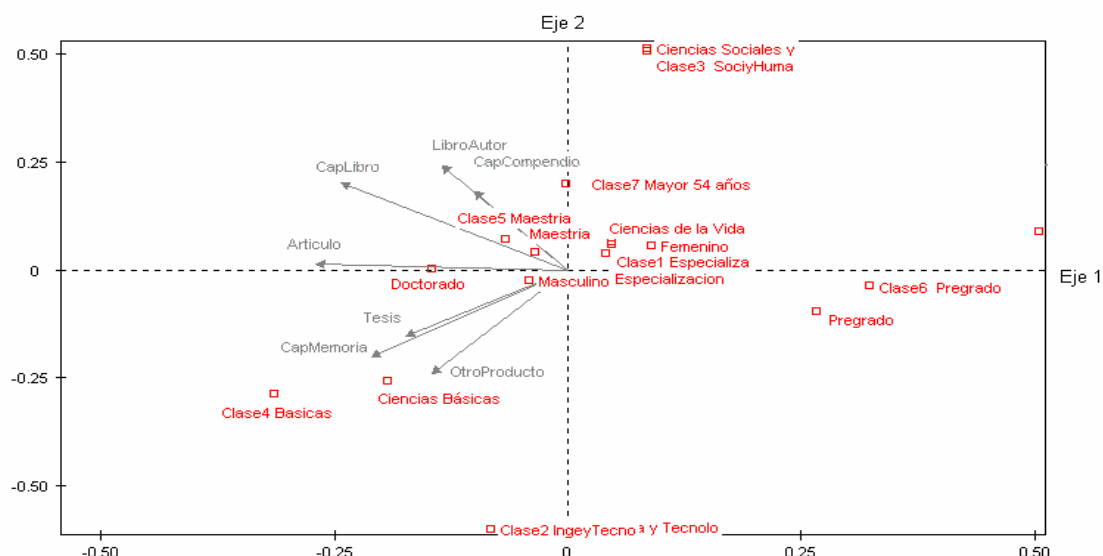


Tabla 1-18: Coordenadas y valores test de las variables nominales ilustrativas

| MODALIDADES | | | | VALORES-TEST | | | | | COORDENADAS | | | | | |
|-----------------------------|------|---------|--|--------------|-------|------|------|------|-------------|-------|-------|-------|-------|--------|
| IDEN - LIBELLE | EFF. | P. ABS | | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | DISTO. |
| 2 . Genero | | | | | | | | | | | | | | |
| Feme - Femenino | 678 | 678.00 | | 2.3 | 1.6 | 2.2 | -1.9 | 0.6 | 0.09 | 0.06 | 0.07 | -0.06 | 0.02 | 0.02 |
| Masc - Masculino | 1470 | 1470.00 | | -2.3 | -1.6 | -2.2 | 1.9 | -0.6 | -0.04 | -0.03 | -0.03 | 0.03 | -0.01 | 0.00 |
| 3 . IdEscolaridad | | | | | | | | | | | | | | |
| Preg - Pregrado | 438 | 438.00 | | 5.0 | -2.1 | -2.7 | 2.7 | 1.0 | 0.27 | -0.10 | -0.12 | 0.11 | 0.04 | 0.11 |
| Espe - Especialización | 212 | 212.00 | | 0.5 | 0.5 | -1.6 | 0.5 | 0.8 | 0.04 | 0.04 | -0.10 | 0.03 | 0.05 | 0.04 |
| Maes - Maestría | 829 | 829.00 | | -1.0 | 1.3 | -1.7 | -0.7 | -2.4 | -0.03 | 0.04 | -0.05 | -0.02 | -0.06 | 0.02 |
| Doct - Doctorado | 669 | 669.00 | | -3.6 | 0.1 | 5.2 | -1.9 | 1.2 | -0.15 | 0.00 | 0.17 | -0.06 | 0.03 | 0.06 |
| 27 . AreaTrabajo_ISI_0 | | | | | | | | | | | | | | |
| ArBe - Arquitectura y Bella | 7 | 7.00 | | 1.1 | 0.2 | -0.7 | 0.2 | -0.2 | 0.50 | 0.09 | -0.25 | 0.07 | -0.07 | 0.36 |
| CBas - Ciencias Básicas | 451 | 451.00 | | -3.7 | -5.6 | 3.9 | -3.2 | 2.1 | -0.19 | -0.26 | 0.17 | -0.13 | 0.08 | 0.17 |
| CSoc - Ciencias Sociales y | 500 | 500.00 | | 1.7 | 12.0 | -5.3 | 2.3 | -2.8 | 0.08 | 0.51 | -0.21 | 0.09 | -0.10 | 0.37 |
| CVid - Ciencias de la Vida | 858 | 858.00 | | 1.4 | 2.3 | 5.6 | -3.4 | 5.2 | 0.05 | 0.07 | 0.15 | -0.09 | 0.13 | 0.07 |
| CInt - Ingeniería y Tecnolo | 332 | 332.00 | | 0.0 | -10.8 | -5.6 | 5.4 | -6.1 | 0.00 | -0.60 | -0.29 | 0.27 | -0.28 | 0.64 |
| 31 . Partition en 7 classes | | | | | | | | | | | | | | |
| Cl1 - Clase1 Especializa | 204 | 204.00 | | 0.6 | 0.8 | -1.3 | 0.2 | 1.0 | 0.05 | 0.06 | -0.08 | 0.02 | 0.06 | 0.03 |
| Cl2 - Clase2 IngeyTecn | 258 | 258.00 | | -1.1 | -9.4 | -4.9 | 4.3 | -5.6 | -0.08 | -0.60 | -0.29 | 0.24 | -0.30 | 0.64 |
| Cl3 - Clase3 SocioyHuma | 330 | 330.00 | | 1.3 | 9.2 | -4.0 | 1.2 | -1.3 | 0.09 | 0.51 | -0.20 | 0.06 | -0.06 | 0.35 |
| Cl4 - Clase4 Básicas | 313 | 313.00 | | -4.8 | -5.0 | 3.5 | -2.8 | 2.9 | -0.31 | -0.29 | 0.18 | -0.14 | 0.14 | 0.28 |
| Cl5 - Clase5 Maestría | 463 | 463.00 | | -1.3 | 1.6 | 5.0 | -2.3 | 1.5 | -0.07 | 0.07 | 0.21 | -0.09 | 0.06 | 0.07 |
| Cl6 - Clase6 Pregrado | 349 | 349.00 | | 5.2 | -0.7 | -0.1 | 0.4 | 1.9 | 0.32 | -0.04 | 0.00 | 0.02 | 0.08 | 0.12 |
| Cl7 - Clase7 Mayor 54 años | 231 | 231.00 | | 0.0 | 2.9 | 0.5 | -0.2 | -1.2 | 0.00 | 0.20 | 0.03 | -0.01 | -0.07 | 0.05 |

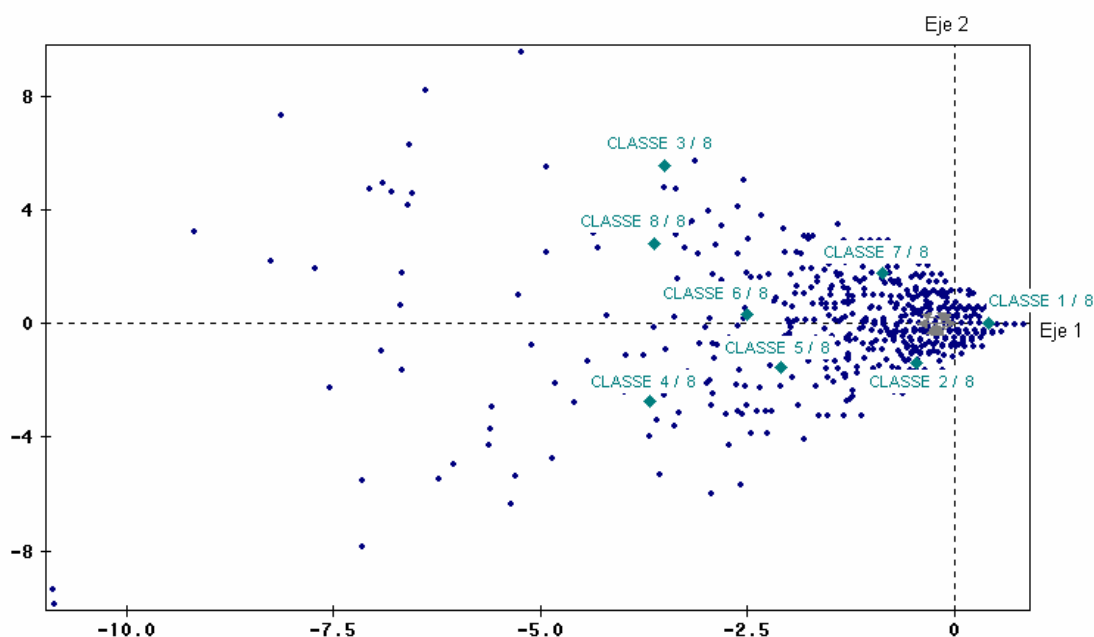
1.3.3. Conclusiones

Como resultado de su actividad los investigadores producen diferentes tipos de resultados. El total de productos logrados por cada uno de los investigadores es un buen indicador de la producción en general. Con el ACP y la proyección de variables ilustrativas, que dividen a los investigadores por otras características se logra una buena síntesis del perfil productivo,. Los investigadores de mayor formación son también los de mayor producción. Los de ciencias básicas son en promedio los más productivos. Se

hacen más tesis en ciencias básicas. Los de ciencias sociales producen más libros y capítulos de libro. Los de ingeniería son los de mayor producción en otros productos.

Dado que las distribuciones son muy asimétricas: unos pocos investigadores producen mucho y muchos producen poco (ver Gráfico 1-14), se puede mejorar el análisis con un análisis de correspondencias múltiples haciendo discretización de las variables de producción por cuantiles (recodificación de cada variable en clases, con frecuencia parecida dentro de cada clase). Con esto se rompe el núcleo denso que aparece en el centro del plano lográndose una mejor descripción.

Gráfico 1-14: investigadores sobre el primer plano factorial



Nota: las clases proyectadas son el resultado de una clasificación que no se presenta aquí. Los puntos representan a los investigadores.

1.4. Ejercicio: test de personalidad

1.4.1. Presentación

A un conjunto de 1142 personas le fueron aplicados dos Test de Personalidad: el PPG-IPG y el 16PF. En la primera prueba los puntajes están en una escala de 0 a 100 y en la segunda de 1 a 10. El primero de ellos es realmente la unión de dos test: el **PPG** (L.D.Gordon) que permite una medida de cuatro aspectos de la personalidad significativos en la vida cotidiana: Ascendencia (ASC), Responsabilidad (RES), Estabilidad Emocional (EST) y Sociabilidad (SOC) y el **IPG** que mide otros cuatro rasgos adicionales: la Cautela CAU), la Originalidad (ORI), la Comprensión (COM) y la Vitalidad (VIT).

Por su parte, el 16PF es otro test de personalidad que consulta 16 aspectos de la personalidad. En este estudio solo se tendrán en cuenta cinco de ellos:

Factor A: (+) Tranquilo – Reservado(-),

Factor C: (+) Ego Fuerte – Ego Débil (-),

Factor E: (+) Dominación – Subordinación (-),

Factor Q2: (+) Autosuficiencia – Suficiente (-),

Factor Q4: (+) Tenso Impulsivo – Relajado (-).

1.4.2. Objetivos

El analista está interesado principalmente en estudiar el Test PPG-IPG y de manera complementaria indagar por el comportamiento de cinco los items del Test 16PF, en el contexto del primero. De una manera mas especifica, el analista está interesado en saber:

- Que tipo de relación existe entre las variables del test PPG-IPG.
- Si existen subconjuntos de variables que están altamente correlacionas.
- Que relación se puede evidenciar entre las variables del Test PPG-IPG y las cinco variables del test 16PF .
- Si el Test PPG-IPG clasifica a los individuos encuestados en clases de individuos diferenciados.

1.4.3. Tablas y gráficos

Se realiza un Análisis de Componentes Principales usando como **Variables Activas** las correspondientes al Test PPG-IPG y como **Variables Ilustrativas** las cinco variables correspondientes al 16PF. A continuación se presentan las y gráficos requeridas para el análisis.

Tabla 1-19: Fragmento.de.la.Base.de.Datos.:Puntajes.en.laPrueba.Sicológica.PPG-IPG y en la Prueba 16PF

| Variables Activas | | | | | | | | | | Variables Ilustrativas | | | | |
|-------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------------------------|---|---|----|----|
| OBS | ASC | RES | EST | SOC | AUT | CAU | ORI | COM | VIT | A | C | E | Q2 | Q4 |
| 1 | 57 | 91 | 73 | 48 | 74 | 48 | 87 | 95 | 43 | 6 | 9 | 5 | 2 | 1 |
| 2 | 29 | 39 | 83 | 27 | 32 | 43 | 71 | 83 | 2 | 7 | 3 | 6 | 8 | 7 |
| 3 | 98 | 39 | 23 | 97 | 85 | 36 | 72 | 92 | 97 | 9 | 6 | 6 | 2 | 5 |
| 4 | 51 | 93 | 83 | 66 | 88 | 70 | 32 | 78 | 64 | 7 | 6 | 7 | 6 | 6 |
| 5 | 92 | 13 | 7 | 96 | 45 | 13 | 97 | 21 | 30 | 6 | 6 | 6 | 9 | 6 |
| 6 | 76 | 85 | 46 | 100 | 100 | 96 | 62 | 23 | 58 | 6 | 8 | 6 | 4 | 3 |
| 7 | 19 | 49 | 57 | 39 | 68 | 70 | 40 | 65 | 64 | 6 | 5 | 5 | 6 | 3 |
| 8 | 79 | 42 | 17 | 85 | 49 | 52 | 63 | 92 | 30 | 7 | 6 | 7 | 8 | 4 |
| 9 | 36 | 97 | 100 | 10 | 84 | 91 | 73 | 60 | 63 | 6 | 5 | 4 | 7 | 5 |
| 10 | 51 | 98 | 95 | 36 | 91 | 95 | 72 | 77 | 42 | 8 | 2 | 4 | 6 | 6 |
| . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 1142 | 17 | 17 | 1 | 52 | 10 | 7 | 30 | 11 | 64 | 4 | 3 | 6 | 7 | 9 |

Tabla 1-20: Estadísticas básicas de las variables continuas del ACP

| Variables Activas | | | | | | | | |
|------------------------|------|------------------------|----------|---------|---------|------------|---------|---------|
| NUM | IDEN | LIBELLE | EFFECTIF | POIDS | MOYENNE | ECART-TYPE | MINIMUM | MAXIMUM |
| 1 | ASC | - Asendencia | 1136 | 1136.00 | 56.31 | 29.22 | 1.00 | 99.00 |
| 2 | RES | - Responsabilidad | 1137 | 1137.00 | 55.23 | 28.72 | 1.00 | 99.00 |
| 3 | EST | - Estabilidad | 1138 | 1138.00 | 57.51 | 27.91 | 1.00 | 100.00 |
| 4 | SOC | - Sociabilidad | 1140 | 1140.00 | 54.02 | 27.46 | 1.00 | 100.00 |
| 6 | CAU | - Cautela | 1138 | 1138.00 | 53.14 | 27.65 | 1.00 | 99.00 |
| 7 | ORI | - Originalidad | 1139 | 1139.00 | 52.92 | 28.72 | 1.00 | 99.00 |
| 8 | COM | - Comprension | 1135 | 1135.00 | 54.12 | 28.13 | 1.00 | 99.00 |
| 9 | VIT | - Vitalidad | 1132 | 1132.00 | 56.55 | 28.23 | 1.00 | 99.00 |
| Variables Ilustrativas | | | | | | | | |
| 10 | A | - Tranquilo-Reservado | 1139 | 1139.00 | 5.74 | 1.98 | 1.00 | 10.00 |
| 12 | C | - EgoFuerte-EgoDebil | 1142 | 1142.00 | 5.64 | 1.81 | 1.00 | 9.00 |
| 13 | E | - Dominacion-Subordina | 1142 | 1142.00 | 5.08 | 1.79 | 1.00 | 10.00 |
| 23 | Q2 | - Autosufic-Suficienci | 1142 | 1142.00 | 5.72 | 1.85 | 1.00 | 10.00 |
| 25 | Q4 | - TensoImpulsivo-Relaj | 1142 | 1142.00 | 5.43 | 1.85 | 1.00 | 10.00 |

Tabla 1-21: matriz de correlaciones

| | ASC | RES | EST | SOC | CAU | ORI | COM | VIT |
|---------------|-------|-------|-------|-------|-------|-------|-------|-------|
| ASC | 1.00 | | | | | | | |
| RES | 0.06 | 1.00 | | | | | | |
| EST | 0.06 | 0.41 | 1.00 | | | | | |
| SOC | 0.52 | 0.01 | -0.22 | 1.00 | | | | |
| CAU | 0.00 | 0.27 | 0.27 | -0.01 | 1.00 | | | |
| ORI | 0.30 | 0.27 | 0.18 | 0.17 | 0.12 | 1.00 | | |
| COM | 0.22 | 0.22 | 0.23 | 0.20 | 0.36 | 0.25 | 1.00 | |
| VIT | 0.19 | 0.17 | 0.12 | 0.16 | 0.06 | 0.21 | 0.03 | 1.00 |
| | ASC | RES | EST | SOC | CAU | ORI | COM | VIT |
| VALORES-TESTS | | | | | | | | |
| | ASC | RES | EST | SOC | CAU | ORI | COM | VIT |
| ASC | 99.99 | | | | | | | |
| RES | 1.97 | 99.99 | | | | | | |
| EST | 1.99 | 14.83 | 99.99 | | | | | |
| SOC | 19.58 | 0.30 | -7.55 | 99.99 | | | | |
| CAU | 0.14 | 9.52 | 9.28 | -0.47 | 99.99 | | | |
| ORI | 10.33 | 9.36 | 6.00 | 5.96 | 4.23 | 99.99 | | |
| COM | 7.69 | 7.70 | 7.74 | 6.80 | 12.57 | 8.51 | 99.99 | |
| VIT | 6.46 | 5.94 | 4.06 | 5.49 | 2.09 | 7.04 | 1.05 | 99.99 |
| | ASC | RES | EST | SOC | CAU | ORI | COM | VIT |

Tabla 1-22: Valores propios del ACP e histograma

| NUMERO | VALEUR PROPRE | POURCENT. | POURCENT. CUMULE | |
|--------|------------------|-----------|---------------------|-------|
| 1 | 2.2527 | 28.16 | 28.16 | ***** |
| 2 | 1.6446 | 20.56 | 48.72 | ***** |
| 3 | 1.0422 | 13.03 | 61.74 | ***** |
| 4 | 0.7956 | 9.95 | 71.69 | ***** |
| 5 | 0.6959 | 8.70 | 80.39 | ***** |
| 6 | 0.6392 | 7.99 | 88.38 | ***** |
| 7 | 0.5607 | 7.01 | 95.39 | ***** |
| 8 | 0.3691 | 4.61 | 100.00 | ***** |

Tabla 1-23: Coordenadas y ejes unitarios para las variables continuas

| Variables Activas | | | | | | | | | | | | | | | | |
|---------------------------|--|-------------|-------|-------|-------|-------|-------------------------------|-------|-------|-------|-------|------------------------|-------|-------|-------|-------|
| VARIABLES | | COORDONNEES | | | | | CORRELATIONS VARIABLE-FACTEUR | | | | | ANCIENS AXES UNITAIRES | | | | |
| IDEN - LIBELLE COURT | | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| ASC - Asendencia | | -0.53 | -0.62 | -0.02 | -0.16 | 0.32 | -0.53 | -0.62 | -0.02 | -0.16 | 0.32 | -0.35 | -0.48 | -0.02 | -0.18 | 0.39 |
| RES - Responsabilidad | | -0.61 | 0.39 | 0.21 | -0.11 | 0.13 | -0.61 | 0.39 | 0.21 | -0.11 | 0.13 | -0.41 | 0.30 | 0.21 | -0.12 | 0.16 |
| EST - Estabilidad | | -0.51 | 0.57 | 0.18 | -0.21 | 0.41 | -0.51 | 0.57 | 0.18 | -0.21 | 0.41 | -0.34 | 0.44 | 0.18 | -0.23 | 0.49 |
| SOC - Sociabilidad | | -0.37 | -0.76 | -0.16 | 0.12 | 0.14 | -0.37 | -0.76 | -0.16 | 0.12 | 0.14 | -0.25 | -0.59 | -0.16 | 0.14 | 0.16 |
| CAU - Cautela | | -0.50 | 0.40 | -0.41 | 0.46 | -0.16 | -0.50 | 0.40 | -0.41 | 0.46 | -0.16 | -0.33 | 0.31 | -0.40 | 0.52 | -0.19 |
| ORI - Originalidad | | -0.62 | -0.13 | 0.20 | -0.42 | -0.60 | -0.62 | -0.13 | 0.20 | -0.42 | -0.60 | -0.41 | -0.10 | 0.19 | -0.47 | -0.71 |
| COM - Comprension | | -0.63 | 0.05 | -0.54 | 0.03 | -0.05 | -0.63 | 0.05 | -0.54 | 0.03 | -0.05 | -0.42 | 0.04 | -0.53 | 0.04 | -0.06 |
| VIT - Vitalidad | | -0.41 | -0.16 | 0.66 | 0.55 | -0.07 | -0.41 | -0.16 | 0.66 | 0.55 | -0.07 | -0.27 | -0.13 | 0.65 | 0.62 | -0.08 |
| Variables Ilustrativas | | | | | | | | | | | | | | | | |
| VARIABLES | | COORDONNEES | | | | | CORRELATIONS VARIABLE-FACTEUR | | | | | ANCIENS AXES UNITAIRES | | | | |
| IDEN - LIBELLE COURT | | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| A - Tranquilo-Reservado | | -0.09 | -0.20 | -0.10 | 0.03 | 0.03 | -0.09 | -0.20 | -0.10 | 0.03 | 0.03 | | | | | |
| C - EgoFuerte-EgoDebil | | -0.16 | 0.01 | -0.02 | -0.01 | 0.01 | -0.16 | 0.01 | -0.02 | -0.01 | 0.01 | | | | | |
| E - Dominacion-Subordina | | 0.04 | -0.29 | 0.13 | -0.13 | -0.05 | 0.04 | -0.29 | 0.13 | -0.13 | -0.05 | | | | | |
| Q2 - Autosufic-Suficienci | | 0.10 | 0.15 | 0.08 | -0.01 | 0.00 | 0.10 | 0.15 | 0.08 | -0.01 | 0.00 | | | | | |
| Q4 - TensoImpulsivo-Relaj | | 0.25 | -0.16 | 0.00 | -0.03 | -0.02 | 0.25 | -0.16 | 0.00 | -0.03 | -0.02 | | | | | |

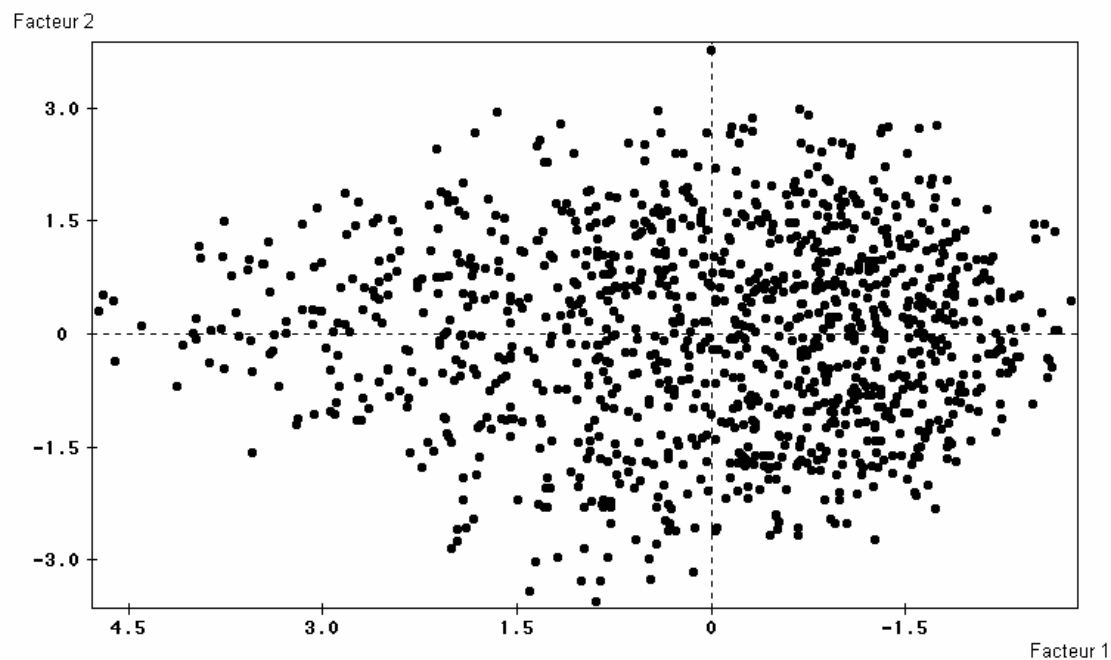
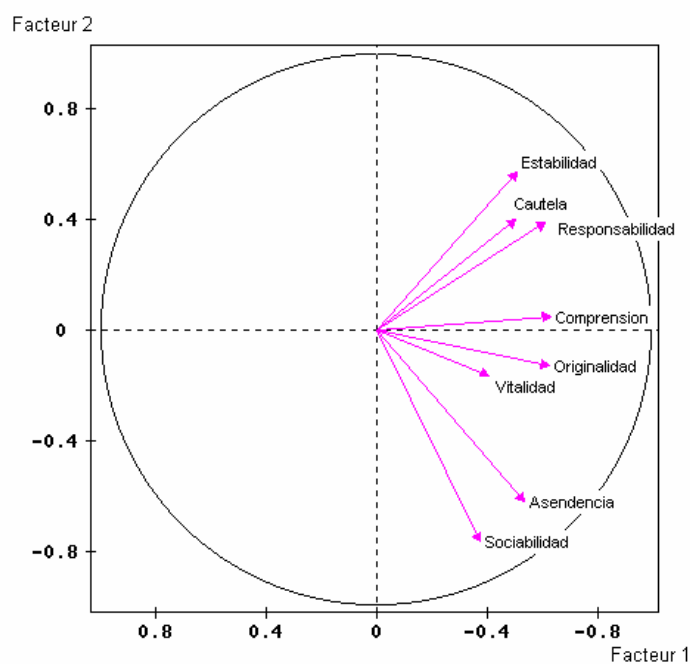
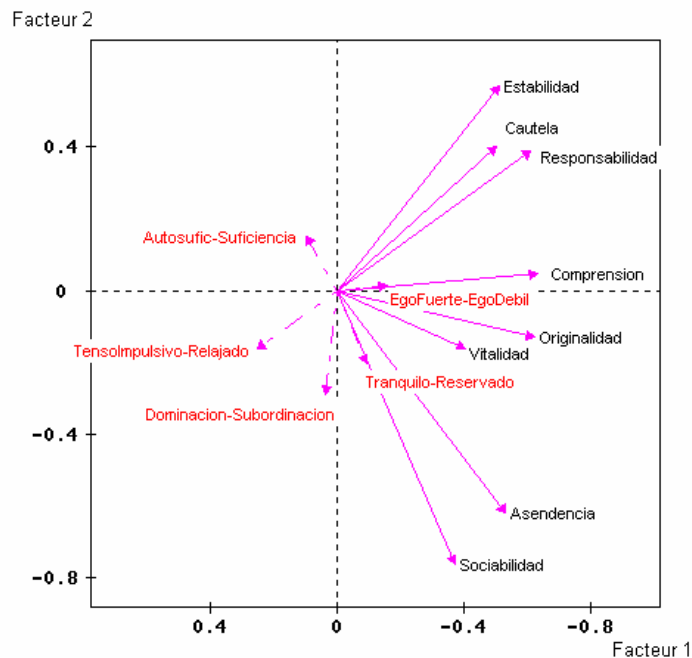
Gráfico 1-15: Primer Plano Factorial. Proyección de los Individuos**Gráfico 1-16: Primer Plano Factorial. Proyección de la Variables Activas**

Gráfico 1-17: Primer plano factorial. proyección de las variables activas, conjuntamente con las Ilustrativas

1.4.4. Guía para el análisis.

1. Porque en este caso es posible hacer un análisis de Componente Principales?
2. Los resultados presentados corresponden a un ACP normado. Es correcto hacerlo en este caso? Podría hacerse un ACP no normado?
3. Que sugieren las mediadas de tendencia central y de variabilidad consignadas en la Tabla 1-20?
4. Según la Tabla 1-21, cuales variables son altamente correlacionadas? Cuales no están correlacionadas?
5. Según el histograma de valores propios (Tabla 1-22), qué porcentaje de la inercia es explicada por los dos primeros ejes factoriales? Cuantos ejes factoriales más se puede sugerir que sean examinados?
6. En la tabla de coordenadas y correlaciones (Tabla 1-23), se observa que todas las coordenadas en el primer factor son negativas. Puede dársele a este hecho alguna interpretación especial?
7. Examinando el Gráfico 1-16:
 - Puede decirse que es coherente con la lectura de la matriz de correlaciones?
 - Que variables se puede decir que están mas altamente correlacionas con el primer factor?
 - Puede darle algún significado a este primer factor?
 - Cuales y en que sentido están altamente correlacionadas con el segundo factor
 - Puede darle algún significado al segundo factor?

- Puede identificar subconjuntos de variables altamente correlacionas entre sí?
 - Existe algún subconjunto de variables que se pueda decir que está no correlacionado con otro subconjunto de variables?
8. Observe simultáneamente el Gráfico 1-15 y el Gráfico 1-16 e identifique puntos individuos que puntúan:
- Alto en la variable Estabilidad y alto en Ascendencia.
 - Alto en Estabilidad pero Bajo en Ascendencia
 - Alto en Comprensión pero bajo en Originalidad.
9. Los análisis anteriores sugieren que pueden constituirse grupos de individuos caracterizados por su puntuación en las distintas variables? Podría sugerir algunos grupos?
10. Al examinar conjuntamente las variables activas e ilustrativas en el Gráfico 1-17 puede decir cuales de estas últimas parecen explicar lo mismo que alguna de las activas o algún subconjunto de ellas?