Análisis espacial

Este tipo de análisis parte de una idea sencilla, aunque su implementación es compleja. Se trata de la ubicación de una unidad de análisis en función de dos o más atributos. Del mismo modo que trazando coordenadas cartesianas podemos ubicar un punto en función de ambos ejes, es posible trazar más ejes y localizar igualmente ese punto. Cada eje es un atributo, por ello Lazarsfeld llamó a esto “espacio de atributos” o “espacio de propiedades”. Si disponemos de dos atributos (por ejemplo edad e ingresos) se nos abre un espacio bidimensional (una superficie). Si agregamos un tercer atributo (por ejemplo, rendimiento escolar), tenemos un espacio tridimensional (volumétrico). Pero si seguimos agregando atributos no lo podremos representar gráficamente, y tenemos lo que se llama un *espacio multidimensional*.

El análisis espacial propone el estudio de una multiplicidad de unidades de análisis en función de una multiplicidad de dimensiones o atributos de los mismos. La vinculación de todos ellos nos permite comprender en profundidad la naturaleza de las poblaciones abordadas. Para ello se establece la matriz de “distancias” o “cercanías” entre las distintas unidades de análisis o entre variables (dimensiones o atributos). De esta manera se pueden correlacionar pares de unidades de análisis, que da lugar a la *matriz Q*, sobre la que se aplican *técnicas Q* de análisis, o bien se pueden correlacionar pares de variables o atributos, sobre las que se origina la *matriz R*, a la que se aplican *técnicas R* de análisis.

Mediante el análisis Q se correlacionan N unidades de análisis en función de k variables, dando como resultado $N(N-1)/2$ pares de distancias entre unidades de análisis. De manera similar procedemos con el análisis R; mediante éste correlacionamos K variables de N unidades de análisis, obteniendo $K(K-1)/2$ pares de distancia entre variables.

Existen distintos tipos de análisis espacial. Presentaremos solamente los cuatro principales, que constituyen los fundamentos de otros tipos de técnicas que no son más que modificaciones de estas principales, ajustándolas a situaciones analíticas particulares.

1) Análisis factorial

El núcleo del análisis factorial (del cual hay muchas variantes) es el principio de parsimonia, o economía de descripción, que es común a todas las ciencias. Este análisis permite determinar *k* variables implícitas o subyacentes partiendo de una serie *n* de medidas, siendo *n* mayor que *k*. Esas variables *k* obtenidas son los llamados "factores", que dan nombre a este tipo de análisis. Como se evidencia, la obtención de un número de factores, menor que el número de medidas de las cuales se toma, simplifica el análisis, siendo ésta la razón de uso de esta técnica.

La mejor forma de comprender el análisis factorial es mediante un ejemplo. Tanto en psicología como en psicopedagogía hay una gran variedad de herramientas de testeo (tests) disponibles, que se pueden utilizar según la preferencia del profesional para chequear aspectos del paciente. Es conocido también que muchos de esos instrumentos abordan aspectos relativamente similares, por ejemplo, relativos a la atención, a la abstracción formal, etc. En función de ello, imaginemos que se toman un conjunto de pruebas a un paciente. Estas pruebas relevan los siguientes aspectos: vocabulario (V), uso de sinónimos (S), capacidad de lectura (L), representación geométrica (G), habilidad aritmética (Ar) y habilidad algebraica (Al). En función de los datos obtenidos, se establecen las correlaciones entre los resultados obtenidos en cada test (que, a nuestros efectos, es una variable).

Eliminando los pares de "gemelos" (el test de vocabulario no se puede comparar con sí mismo, porque carece de sentido), el resultado de la correlación entre cada par de variables es el siguiente:

V–S: 0,85	V–L: 0,59	V–G: 0,42	V–Ar: 0,37	V–Al: 0,16
S–L: 0,61	S–G: 0,28	S–Ar: 0,03	S–Al: 0,39	L–G: 0,12
L–Ar: 0,33	L–Al: 0,44	G–Ar: 0,70	G–Al: 0,63	Ar–Al: 0,91

Con estos de correlación datos obtenidos se construye la matriz *R*. (Recordemos que, por tratarse de correlación entre variables *k* utilizamos este tipo de técnica. Si fuese entre unidades de análisis, construiríamos una matriz *Q*).

Matriz *R*

	V	S	L	G	Ar	Al
V	—	0,85	0,59	0,42	0,37	0,16
S	0,85	—	0,61	0,28	0,03	0,39
L	0,59	0,61	—	0,12	0,33	0,44
G	0,42	0,28	0,12	—	0,70	0,63
Ar	0,37	0,03	0,33	0,70	—	0,91
Al	0,16	0,39	0,44	0,63	0,91	—

Puestos los datos en la matriz *R* —y considerando como correlación “alta” aquella cuyos valores están por encima de 0,50; y como correlación “baja” aquellos que no superan 0,50— es fácil observar la configuración de dos conglomerados:

Conglomerado I

	V	S	L	G	Ar	Al
V	—	0,85	0,59	0,42	0,37	0,16
S	0,85	—	0,61	0,28	0,03	0,39
L	0,59	0,61	—	0,12	0,33	0,44
G	0,42	0,28	0,12	—	0,70	0,63
Ar	0,37	0,03	0,33	0,70	—	0,91
Al	0,16	0,39	0,44	0,63	0,91	—

Conglomerado II

El análisis factorial nos ha agrupado dos conglomerados, el I, que agrupa las cuestiones relativas al vocabulario, los sinónimos y la lectura; y el II, que agrupa las destrezas en geometría, aritmética y álgebra. Es decir que nos ha reducido las seis variables iniciales a dos, mostrando que existe una alta correlación entre lo que miden los tests de vocabulario, uso de sinónimos y lectura (correlaciones que varían entre 0,59 y 0,85) por un lado, y los que miden la habilidad en geometría, aritmética y álgebra (correlacionados entre 0,63 y 0,91) por otro.

De esta manera el análisis factorial nos ha permitido hacer más sencillo el examen de la situación. Lo que no brinda el análisis factorial —y queda a cargo del analista— es la significación de tales conglomerados. En el ejemplo que hemos puesto, el conglomerado I podría denominarse “Expresión” y el conglomerado II “Matemático”, para dar una idea de los factores que aparecen.

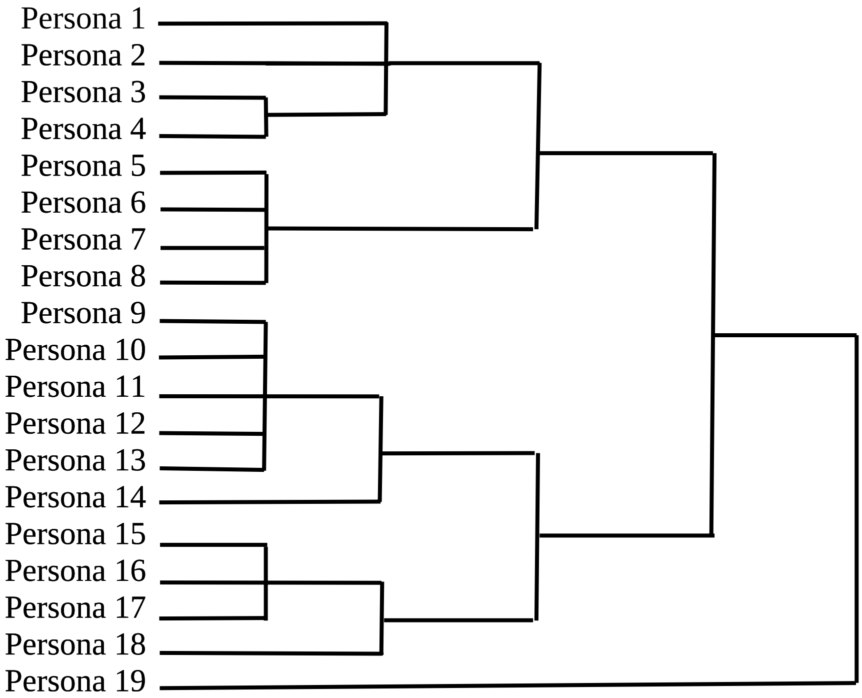
Por supuesto, esta presentación es sumamente sencilla. En la práctica real suelen utilizarse muchas más variables, y los conglomerados no necesariamente se agrupan en bloque como en este ejemplo. Suelen aparecer bloques y celdas sueltas que pertenecen al mismo conglomerado, dificultando a veces su interpretación. Vale la pena recalcar una vez más que para esto es insustituible la teoría.

En este ejemplo hemos presentado un análisis factorial de un solo paso. En general se suelen realizar varios pasos (los factores obtenidos vuelven a considerarse variables y así se obtienen nuevos factores). Cuando una variable es idéntica al factor resultante (mide eso y solo eso) se dice que es factorialmente “pura” y que, por lo tanto, está *saturada* por el factor. Si, por el contrario, está saturada por varios factores (es decir que estos factores se vuelven irreductibles) se dice que es factorialmente “compleja”. En realidad, el análisis factorial se termina cuando, después de la cantidad de pasos necesarios, puede establecerse una variable factorialmente pura o se satura; es decir, cuando volviendo a correlacionar todos los pares existentes, no nos arroja como resultado una mayor simpleza, sino que todo sigue igual.

2) Análisis de conglomerados (“Cluster analysis”)

El análisis de conglomerados se utiliza para ver la forma en que las unidades de análisis se una población de estudio se asemejan o difieren entre sí. Difiere del análisis factorial en el tipo de organización que se hace de los datos. Por empezar, es un tipo de análisis *Q*, aquí no se consideran las variables (análisis *R*). El tipo de organización se grafica con un dendograma (que presentaremos más abajo), y se pueden utilizar variables cuantitativas y cualitativas.

Imaginemos que consideramos una población laboral de una empresa, de la que observamos atributos tales como: edad, ingresos, sexo, antigüedad, estado civil, rendimiento laboral, grado de satisfacción con la tarea desempeñada, etc. En función de los datos obtenidos se miden las distancias entre ellos (adviértase la dificultad que representa medir distancias entre todas las unidades de análisis estableciendo pares, tríos, cuartetos, etc., de acuerdo a su grado de cercanía o lejanía de otros conglomerados o “clusters”, tanto más cuanto que estas distancias se toman con atributos de diferente nivel de medición). Obtenidos estas distancias, se organizan de la siguiente manera:



Este esquema, llamado *dendograma*, articula conglomerados, en distintos niveles. El análisis finaliza en el último nivel de agregación. Pero cada nivel que desagregamos (leyéndolo de derecha a izquierda) nos aporta información. En general se suele considerar el último nivel o a lo sumo el anteúltimo. Aquí tenemos que en el último nivel se articulan la persona 19 con el resto del conjunto, y en el anteúltimo tenemos tres conglomerados o *clusters*, el formado por las personas 1–8, el formado por las personas 9–18 y el formado por la persona 19.

De este dendograma se establece fácilmente que la persona 19 es completamente diferente a las demás (habrá que determinar si por particularidades positivas o negativas, pero eso es otro tipo de análisis), en tanto 1, 2, 14 y 18, aunque agrupables, no lo son en una primera instancia (obsérvese que se las puede agrupar en una segunda "corrida" o segundo paso del análisis de conglomerados.

3) Análisis de segmentación ("Tree analysis")

Un tercer tipo de análisis espacial es el de segmentación. Se lo conoce también como análisis arborescente (dado que forma un árbol invertido) y es, en su lógica, el opuesto al recién visto, ya que va desagregando (segmentando) la población de estudio, tanto como es posible. La segmentación se realiza mediante dicotomizaciones en cada variable considerada.

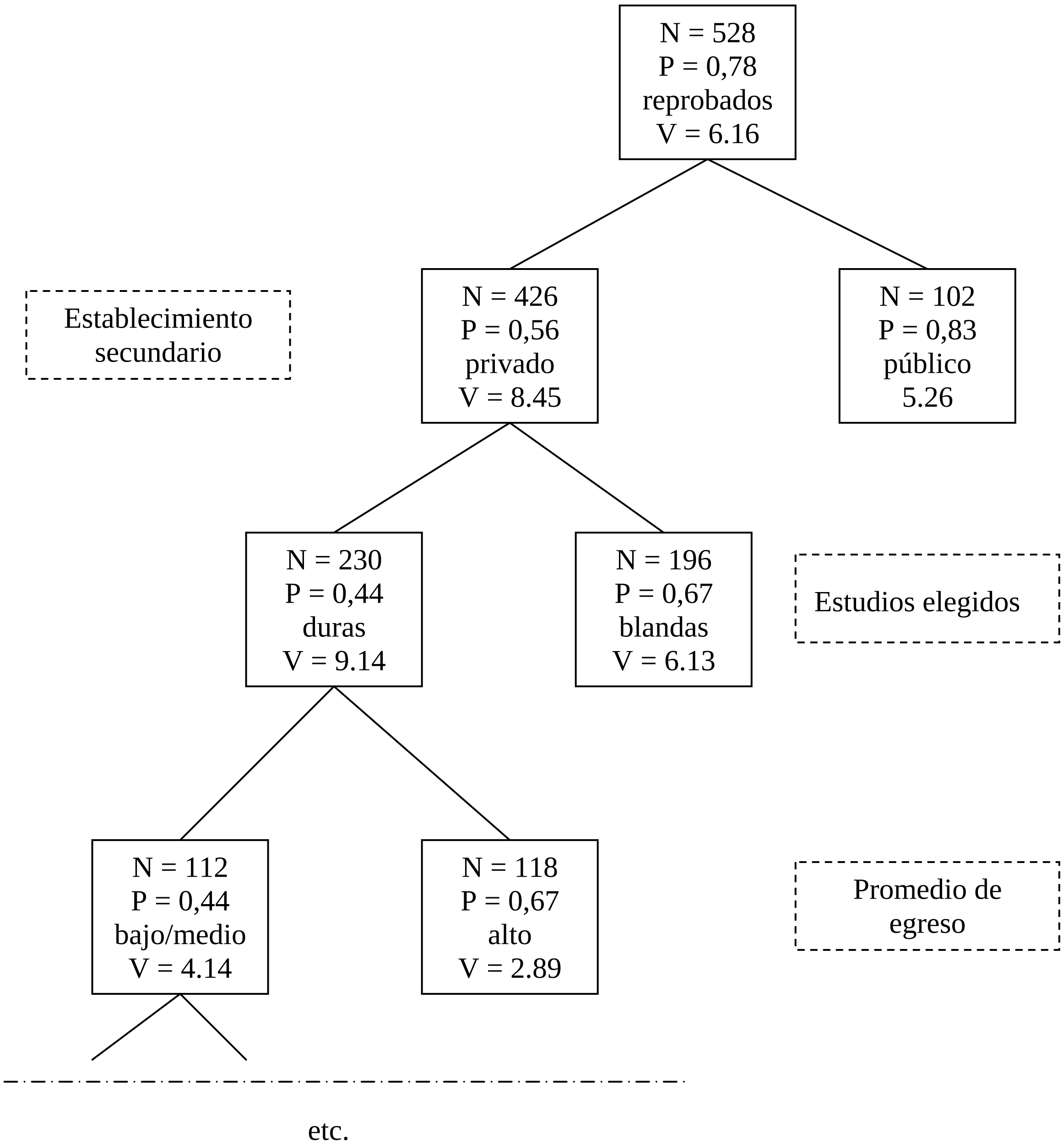
La operatoria es la siguiente: de todas las variables que contiene nuestra población se observa la varianza obtenida —también puede observarse el desvío típico—; aquella variable cuya dicotomización —es decir, su división en dos— produzca los dos grupos más homogéneos —al menos uno de ellos debe serlo— será considerada “dependiente” a los efectos de ser dividida, mientras el resto de las variables siguen siendo consideradas “independientes”. Recordemos que la varianza —o el desvío típico— es una medida de dispersión: una curva bimodal, por ejemplo, tiene una alta dispersión (medida con la varianza o con el desvío típico) y su dicotomización conformará dos grupos relativamente homogéneos. De esta manera, el comportamiento de esta variable “dependiente” se explica por el conjunto de las variables “independientes” restantes.

Realizado este paso, se vuelve a considerar al resto de las variables, hasta encontrar aquella cuya varianza determine que su dicotomización conformará dos grupos relativamente homogéneos, y diferenciados del resto de las demás variables. Nuevamente se ha buscado la mejor forma de explicación del comportamiento de una variable. Así se continúa hasta el agotamiento de las variables o bien hasta que —aún cuando queden variables— la varianza adicional ya no resulte significativa.

Veamos un ejemplo. En una universidad la tasa de reprobación del examen de ingreso se ha incrementado de un 35% a un 78% en seis años. Dada la magnitud del fenómeno se intentan establecer las causas del fracaso, y cuáles son las variables estructurales operantes en el mismo. Para ello se recurre a un análisis de segmentación sobre aquellas variables que teóricamente se consideran relevantes: establecimiento secundario de origen (estatal o privado), orientación elegida (ciencias duras o blandas), promedio de egreso (bajo/medio o alto), estudios de los padres (básicos/medios o superiores), existencia de hermanos/as en el nivel superior/universitario (tiene o no), radicación geográfica del estudiante (lugareño o migrante).

Se realizan los cálculos de varianza para cada variable y a partir de allí se comienza el análisis de segmentación. Para cada variable se establece, además P (la probabilidad de encontrar un caso favorable a uno de los términos de la dicotomía, mientras Q es la probabilidad de encontrar un caso negativo).

Lo que obtenemos es lo siguiente:



De esta manera se ha ido segmentando la población en distintos estratos, cada uno de los cuales explica el comportamiento de una porción de la misma, según las distintas variables dependientes que hemos ido considerando.

Es necesario advertir que, excepcionalmente, se pueden dicotomizar los dos grupos de una variable, para dar final al análisis. De esta manera podríamos llegar a un final con cuatro grupos y no solamente dos.