

# Análisis multivariante

Teresa Villagarcía

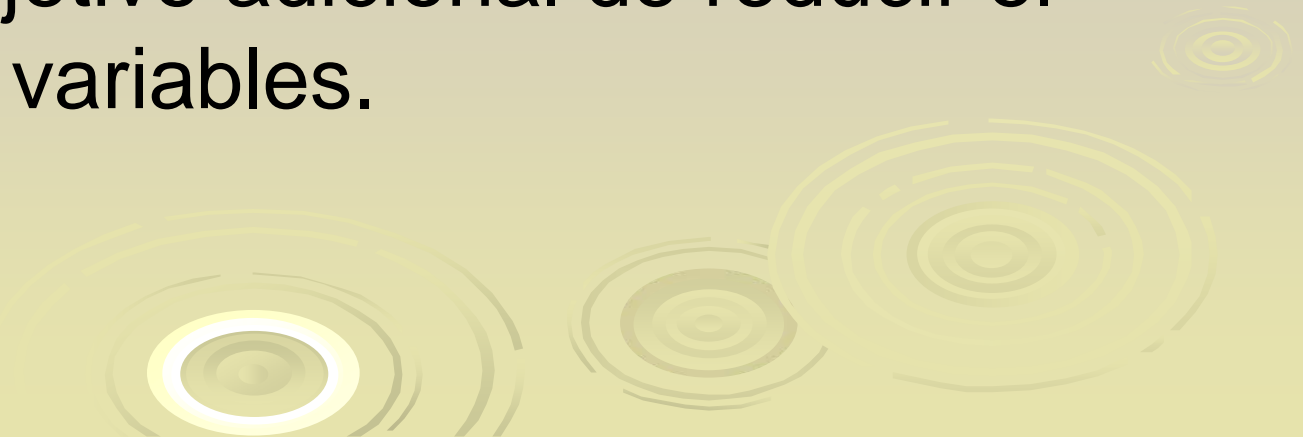


# Análisis de datos


- Nos encontramos grandes masas de datos.
- Hay que estudiarlas con herramientas informáticas.
- Por ejemplo:
  - Encuestas con muchas observaciones y muchas variables.
  - Extraer información es muy laborioso.

## El análisis multivariante

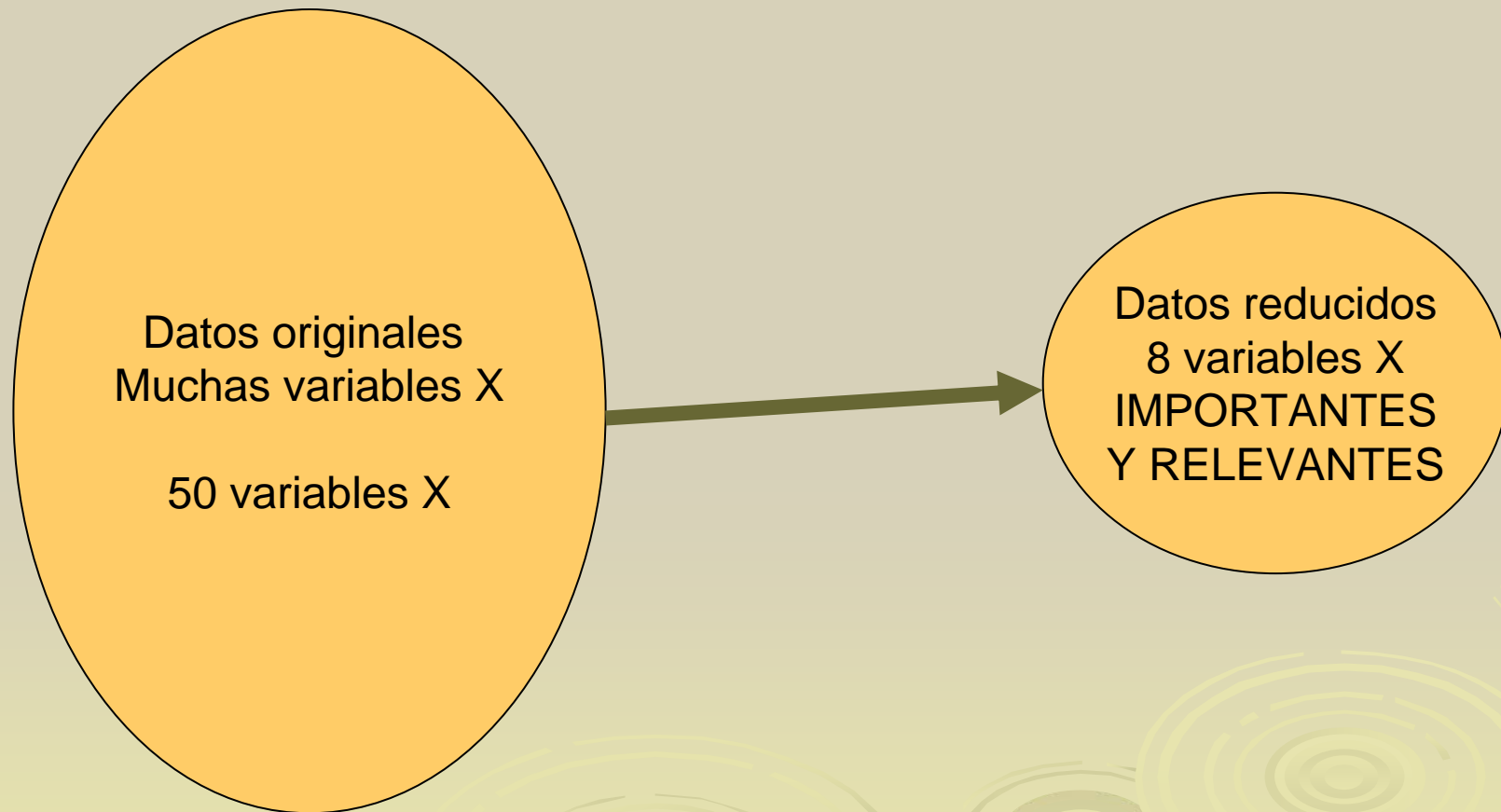
- Sirve para analizar grandes masa de datos
- Utiliza herramientas informáticas
- Extrae la información contenida en los datos
- Tiene el objetivo adicional de reducir el número de variables.



## Vamos a estudiar:

- Análisis Factorial
  - Análisis Cluster
  - Aplicación de todo lo visto a satisfacción de clientes:
    - Regresión
    - Factorial
    - Regresión
  - Obtención de pesos
- 

# Análisis Factorial



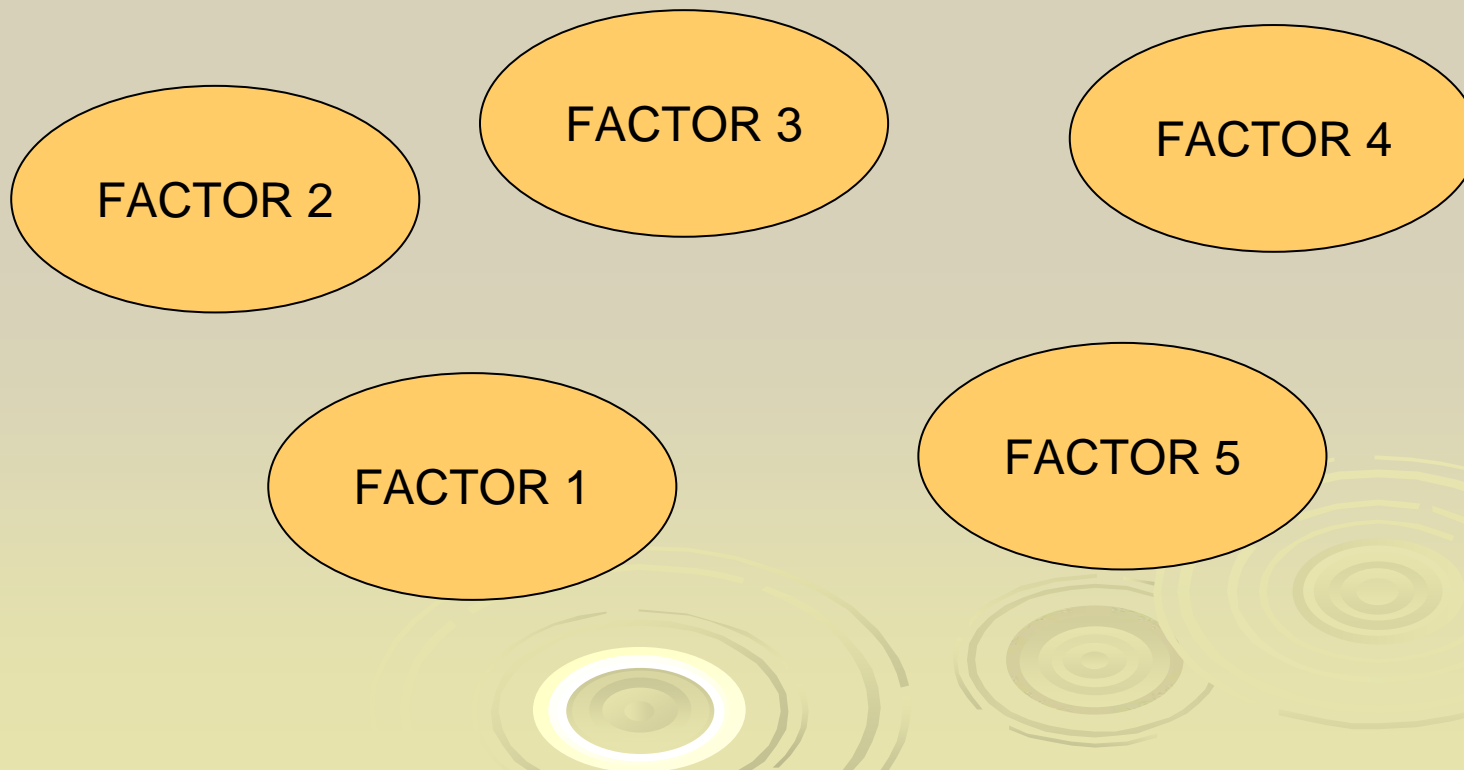
# Análisis factorial

- Tenemos muchas variables con información común. Es decir muy relacionadas
- Creemos que las relaciones existen porque las variables son consecuencia de la existencia de unos factores no observables directamente
- Se pretende llegar a un cálculo de esos factores. Así conseguiremos resumir la información y entender la relación entre las variables
- En el futuro podremos reducir el número de variables en el cuestionario



# Concepto de Factor

- Pensamos que hay unos cuantos “Factores”, pocos que son los determinantes de nuestras opiniones



# Concepto de Factor

El factor es nuestra visión interna y muchas variables tendrán unos valores determinados si el factor es de una manera determinada.

Ejemplo:

Ser católico del Opus Dei implica que si te preguntan:

¿Estás a favor del aborto?

¿Estás a favor de que exista el divorcio?

¿Admites el matrimonio homosexual?

Vas a contestar que no.

Ejemplo:

Ser muy inteligente en matemáticas implica:

Tienes facilidad de cálculo

Comprendes problemas y lenguajes matemáticos

Tienes buena lógica.



Inteligencia verbal  
alta

Inteligencia matemática  
normal



Notas muy buenas en: lectura, escritura, se expresa bien,  
entiende bien lo que se dice.....

.....Y todo lo que tenga que ver con que tiene muy buena habilidad verbal

Notas normales en: Cálculo, Geometría, No es muy buena en relaciones numéricas  
.....Y todo lo que tenga que ver con las habilidades matemáticas

# Origen del AF

- El origen del AF proviene del intento de los psicólogos de entender los factores determinantes de la inteligencia humana.
- Las variables observadas  $X$  eran los resultados de los diferentes tests de inteligencia.
- Se pensaba que estos resultados dependían muy directamente de unos Factores desconocidos que hacían que una persona fuera más hábil en determinadas pruebas.
- Los factores resultaron ser lo que posteriormente se denominó Inteligencia Verbal, Matemática, Espacial etc.
- Así, los resultados que se obtenían en la pruebas de matemáticas estaban muy determinados por la Inteligencia Matemática.

# Ejemplo: Medir la inteligencia de un grupo de alumnos de Administración de Empresas

## ➤ Variables observadas (X)

- Nota de Estadística
- Nota de Matemáticas
- Nota de Derecho
- Nota de Economía general
- Nota de Econometría
- Nota de contabilidad
- Nota de Derecho privado
- Nota de Contabilidad
- Nota de Economía de la Empresa

## ➤ Se identifican tres Factores:

- Capacidad matemática
- Capacidad de memoria
- Capacidad analítica

## El Modelo.

$$X_1 = a_{11}F_1 + a_{12}F_2 + \cdots + a_{1k}F_k + U_1$$

$$X_2 = a_{21}F_1 + a_{22}F_2 + \cdots + a_{2k}F_k + U_2$$

...

$$X_p = a_{p1}F_1 + a_{p2}F_2 + \cdots + a_{pk}F_k + U_p$$

Es decir, el valor de las variables depende del nivel de los factores

## La información de las variables depende:

$$\begin{aligned} X_1 &= a_{11}F_1 + a_{12}F_2 + \cdots + a_{1k}F_k + U_1 \\ X_2 &= a_{21}F_1 + a_{22}F_2 + \cdots + a_{2k}F_k + U_2 \\ &\quad \dots \\ X_p &= a_{p1}F_1 + a_{p2}F_2 + \cdots + a_{pk}F_k + U_p \end{aligned}$$

Depende de los factores  
COMUNALIDAD

Y de algo más, desconocido y  
específico de cada variable

## Procedimiento:

1. Seleccionar las variables que queremos reducir
2. Seleccionar el número de factores adecuado:
  1. Método de los autovalores (No muy bueno)
  2. Pro puro sentido común (Mucho mejor)Se utiliza el porcentaje de varianza explicado
3. Rotar los factores para obtener un resultado comprensible

## Lo vamos a estudiar con un ejemplo

- La cata de pescados.
- Las variables de cata estaban muy relacionadas
- Había mucha colinealidad
- Tras muchos análisis llegamos a la conclusión de que hay variables que sobran



# Cata de pescados

## Multiple Regression Analysis

Dependent variable: ACEPTAGENE

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	3,23662	0,493396	6,55987	0,0000
ACUOSIDAD	0,424113	0,0860697	4,92755	0,0000

## Multiple Regression Analysis

Dependent variable: ACEPTAGENE

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	2,25431	0,338362	6,66241	0,0000
JUGOSIDAD	0,681838	0,0647102	10,5368	0,0000

Colinealidad

## Multiple Regression Analysis

Dependent variable: ACEPTAGENE

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	2,17245	0,414468	5,24154	0,0000
JUGOSIDAD	0,666774	0,0782931	8,51639	0,0000
ACUOSIDAD	0,0286251	0,0831194	0,344386	0,7311



# Mas variables

Multiple Regression Analysis				
Dependent variable: ACCEPTAGENE				
Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	2,33126	0,536478	4,34549	0,0000
ASPECTGENE	0,497581	0,07981	6,23458	0,0000

Colinealidad

Multiple Regression Analysis				
Dependent variable: ACCEPTAGENE				
Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	1,82866	0,44068	4,14962	0,0001
COLOR	0,589448	0,0664088	8,87604	0,0000

Multiple Regression Analysis				
Dependent variable: ACCEPTAGENE				
Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	1,94446	0,488237	3,98262	0,0001
COLOR	0,642279	0,1158	5,54643	0,0000
ASPECTGENE	-0,0697154	0,125018	-0,557644	0,5781

# Ejemplo: Variables de cata de los pescados

Cada catador evalúa diversos aspectos de la cata:  
Aspecto General, Color, Olor, Acuosiad, Jugosidad,  
Dureza, Firmeza y Flavor.

Se introducen las variables al ordenador y se obtiene la  
TABLA DE AUTOVALORES y varianzas explicadas



# Ejemplo: Variables de cata de los pescados

Se introducen al programa las variables y se analiza la TABLA DE AUTOVALORES con los porcentajes de varianza explicada

## Factor Analysis

Factor Number	Eigenvalue	Percent of Variance	Cumulative Percentage
1	3,60352	45,044	45,044
2	1,71301	21,413	66,457
3	1,01183	12,648	79,105
4	0,51970	6,496	85,601
5	0,43689	5,461	91,062
6	0,32336	4,042	95,104
7	0,25211	3,151	98,256
8	0,13955	1,744	100,000

# Elección del número de factores

## Factor Analysis

Factor Number	Eigenvalue	Percent of Variance	Cumulative Percentage
1	3,60352	45,044	45,044
2	1,71301	21,413	66,457
3	1,01183	12,648	79,105
4	0,51970	6,496	85,601
5	0,43689	5,461	91,062
6	0,32336	4,042	95,104
7	0,25211	3,151	98,256
8	0,13955	1,744	100,000

- Si reducimos la 8 variables a un factor, el factor contendría el 45% de la información original de las variables
- Con dos factores el 66.45%
- Con tres factores el 79,1%

- **HAY QUE ELEGIR QUE CANTIDAD DE INFORMACION PERDIDA NOS PARECE ACEPTABLE**

3, 4 o 5 factores

# Tres Factores 79%

## Composición de los factores rotados

Factor Loading Matrix After Varimax Rotation

	Factor 1	Factor 2	Factor 3
ASPECTGENE	0,904446	0,0528491	0,0756087
COLOR	0,938067	0,156675	0,0903128
OLOR	0,547074	0,462024	0,408179
FLAVOR	0,664154	0,295072	0,424016
ACUOSIDAD	0,0502852	-0,126084	0,872586
FIRMEZA	0,237533	0,877522	0,0196523
JUGOSIDAD	0,224296	0,0407906	0,828086
DUREZA	0,0612992	0,917944	-0,10524

**¿Nos aclara algo?**

- Primer Factor: Aspecto General, Color, Flavor
- Segundo Factor: Olor, Firmeza y dureza
- Tercer Factor: Olor Flavor Jugosidad y Acuosidad

# Tres Factores 79%

## Composición de los factores rotados

Variable	Estimated Communality
-----	
ASPECTGENE	0,826533
COLOR	0,912673
OLOR	0,679367
FLAVOR	0,707958
ACUOSIDAD	0,779832
FIRMEZA	0,826854
JUGOSIDAD	0,737698
DUREZA	0,857455
-----	

Las variables quedan explicadas por los factores al porcentaje indicado

## 4 Factores: 85.6%

### Composición de los factores rotados

Factor Loading Matrix After Varimax Rotation

	Factor 1	Factor 2	Factor 3	Factor 4
ASPECTGENE	0,939077	0,117426	0,161682	0,148063
COLOR	0,834063	0,128938	0,446158	0,0584748
OLOR	0,220667	0,281357	0,831219	0,202614
FLAVOR	0,383018	0,146853	0,760841	0,255328
ACUOSIDAD	0,044415	-0,103478	0,0889894	0,896268
FIRMEZA	0,204923	0,888258	0,188476	0,0298655
JUGOSIDAD	0,15469	0,0288791	0,262614	0,812616
DUREZA	0,0237076	0,91514	0,147937	-0,110069

- Primer Factor: Aspecto General y Color
- Segundo Factor: Firmeza y Dureza
- Tercer Factor: Olor y Flavor (sabor)
- Cuarto Factor: Acuosidad y Jugosidad

**¿Nos aclara algo?**

## 4 Factores: 85.6%

### Composición de los factores rotados

Variable	Estimated Communality
-----	
ASPECTGENE	0,943718
COLOR	0,914761
OLOR	0,859833
FLAVOR	0,812339
ACUOSIDAD	0,823896
FIRMEZA	0,867412
JUGOSIDAD	0,754074
DUREZA	0,872045
-----	

Las variables quedan explicadas por los factores al porcentaje indicado



# 5 Factores: 91%

¿Nos aclara algo?

## Composición de los factores rotados

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
ASPECTGENE	0,940625	0,1173	0,171514	0,115822	0,0790595
COLOR	0,830945	0,124247	0,439729	-0,0476496	0,125671
OLOR	0,218786	0,27581	0,83222	0,0765675	0,193606
FLAVOR	0,385091	0,145823	0,77642	0,196724	0,129743
ACUOSIDAD	0,0608836	-0,0978239	0,156246	0,94124	0,260703
FIRMEZA	0,20207	0,884248	0,182311	-0,0462499	0,109695
JUGOSIDAD	0,144478	-0,00153784	0,230726	0,290799	0,912023
DUREZA	0,0242781	0,918696	0,153382	-0,0572071	-0,0929981

- Primer factor: Aspecto general y Color
- Segundo factor: Firmeza y Dureza
- Tercer factor: Olor y Flavor
- Cuarto factor: Acuosidad
- Quinto factor: Jugosidad
- Empieza a separar las variables una por una

# 5 Factores: 91%

## Composición de los factores rotados

Variable	Estimated Communality
-----	
ASPECTGENE	0,947617
COLOR	0,917331
OLOR	0,859875
FLAVOR	0,827922
ACUOSIDAD	0,991587
FIRMEZA	0,870136
JUGOSIDAD	0,990462
DUREZA	0,880039
-----	

Las variables quedan explicadas por los factores al porcentaje indicado

# ¿Qué escogeríamos?

- 4 factores:
  - Aspecto externo
  - Textura al masticar
  - Boca
  - Jugosidad
- Cualquier variable que añadamos que tenga que ver con los 4 factores dará valores muy correlados con las anteriores.

- Primer Factor: Aspecto General y Color
- Segundo Factor: Firmeza y Dureza
- Tercer Factor: Olor y Flavor (sabor)
- Cuarto Factor: Acuosidad y Jugosidad

# Ejemplo: Ingeniería Técnica Mecánica

## ➤ Atributos parciales de calidad:

- Organiza clases
- Entusiasmo
- Participación
- Puntualidad
- Despacho
- Lecturas
- Prácticas pizarra
- Prácticas ordenador

# Ingeniería Mecánica

## Factor Analysis

Factor Number	Eigenvalue	Percent of Variance	Cumulative Percentage
1	4,49655	56,207	56,207
2	1,07716	13,464	69,671
3	0,803097	10,039	79,710
4	0,574503	7,181	86,891
5	0,360583	4,507	91,399
6	0,32942	4,118	95,516
7	0,221527	2,769	98,285
8	0,137165	1,715	100,000

¿Cuántos?

# Ingeniería Mecánica: 3

	Factor 1	Factor 2	Factor 3
	-----	-----	-----
Organiza_clases	0,667032	0,484441	0,267565
Entusiasmo_interes	0,744223	0,430872	0,325457
Prom_participacion	0,870207	0,303578	0,0488701
Lec_Bibliografia	0,683836	-0,187494	0,470227
Puntualidad	0,0932509	0,429551	0,775702
Dispon_consultas	0,351196	0,175744	0,803601
Clase_Practica	0,182963	0,863114	0,133808
Sati_claspractica	0,235795	0,815134	0,305184

# Ingeniería Mecánica: 4

	Factor 1	Factor 2	Factor 3	Factor 4
Organiza_clases	0,783814	0,283534	0,344863	0,0992807
Entusiasmo_interes	0,820586	0,244298	0,371986	0,193353
Prom_participacion	0,838534	0,220115	0,0123268	0,321357
Lec_Bibliografia	0,271912	0,0510789	0,142547	0,915326
Puntualidad	0,224781	0,25582	0,866579	0,0288013
Dispon_consultas	0,227923	0,186183	0,696874	0,480459
Clase_Practica	0,225363	0,912535	0,135083	0,0548853
Sati_claspractica	0,294361	0,810235	0,321619	0,0841255

# Ingeniería Mecánica: 5

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Organiza_clases	0,795297	0,284211	0,0355878	0,185547	0,315189
Entusiasmo_interes	0,822667	0,245044	0,188129	0,306913	0,209714
Prom_participacion	0,83277	0,220161	0,344009	0,0182654	0,0107154
Lec_Bibliografia	0,261538	0,0517605	0,917433	0,0890827	0,214467
Puntualidad	0,221013	0,257566	0,0925268	0,901482	0,238408
Dispon_consultas	0,254921	0,187904	0,285172	0,277122	0,833214
Clase_Practica	0,222721	0,912796	0,077355	0,155007	0,0158835
Sati_claspractica	0,304178	0,810922	0,0267479	0,18575	0,285308



# ¿Con qué nos quedamos?

➤ Factor 1:

.....

