INFERENCIA Y ANÁLISIS FACTORIAL

John Wishart (1898, 1956): Estadístico británico. Fue ayudante de Pearson y Fisher. Encontró una distribución multivariante que lleva su nombre y contribuyó al diseño de experimentos. Profesor de Estadística y Agricultura en Cambridge, fue también consultor de la FAO. Visitó España en 1947.

Prof. ESPERANZA AYUGA TÉLLEZ

INFERENCIA

El problema central en el análisis de datos es decidir si las propiedades encontradas en la muestra se pueden generalizar a la población de la que proviene⇒construir un modelo.

En el análisis multivariante, al aumentar la dimensión aumenta la complejidad (espacio + vacío).

Son necesarias muestras de $n>10p$ y aconsejables $n>20p$.

DISTRIBUCIONES

❋ Multinomial: extensión de la binomial.

Si de una población dividida en G grupos, con p_i probabilidad de que un elemento pertenezca a i

$\sum_{i=1}^{G} p_i = 1$ tomamos n individuos $\sum_{i=1}^{G} n_i = n$:

$$P(x_1 = n_1, \ldots, x_G = n_G) = \frac{n!}{n_1! \ldots n_G!} p_1^{n_1} \ldots p_G^{n_G}$$

$$E(\mathbf{X}) = n\mathbf{P} = \boldsymbol{\mu} \qquad V(\mathbf{X}) = n\left[\mathrm{diag}(\mathbf{P}) - \mathbf{P}\mathbf{P}'\right]$$

DISTRIBUCIONES

* **Dirichlet**: para $0 \leq x_i \leq 1$, $\Sigma x_i = 1$ (las x_i no independ.)

$$f(x_1,\ldots,x_G) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)\ldots\Gamma(\alpha_G)} x_1^{\alpha_1 - 1} \ldots x_G^{\alpha_G - 1}$$

Con $\alpha = (\alpha_1,\ldots,\alpha_G)'$ vector de parámetros de la distrib.

$\alpha_0 = \alpha'\mathbf{1} = \Sigma\alpha_i$

$$V(\mathbf{X}) = \frac{1}{\alpha_0 + 1}\left[(1/\alpha_0)\mathrm{diag}(\boldsymbol{\alpha}) - \boldsymbol{\alpha}\boldsymbol{\alpha}'\right]$$

$$E(\mathbf{X}) = \boldsymbol{\alpha} / \alpha_0$$

$$\mathrm{Cov}(x_i x_j) = \frac{\alpha_i \alpha_j}{\alpha_0^2(\alpha_0 + 1)}$$

DISTRIBUCIONES

- Normal k-dimensional: extensión de la normal.

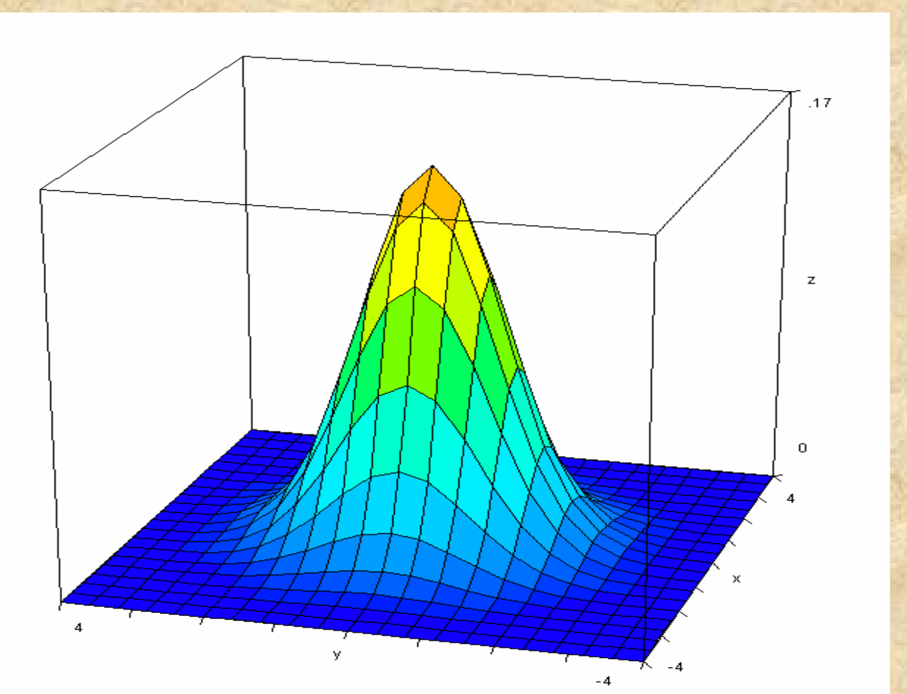
$$f(\mathbf{X}) = \left|\mathbf{V}\right|^{-1/2} (2\pi)^{-p/2} \exp\{-(1/2)(\mathbf{X}-\boldsymbol{\mu})' \mathbf{V}^{-1} (\mathbf{X}-\boldsymbol{\mu})\}$$

$$E(\mathbf{X}) = \boldsymbol{\mu} \qquad V(\mathbf{X}) = \mathbf{V}$$

Al cortar por hiperplanos paralelos a los definidos por las p-var. se obtienen las curvas de nivel (elipses):

$(X-\mu)'V^{-1}(X-\mu)=cte=D^2$

D=dist. de Mahalanobis



DISTRIBUCIONES

❋ de Wishart: es la distribución muestral de las matrices de covarianzas en muestras de normales multivariantes. Los parámetros son m (g.d.l.) y Σ matriz de covarianzas.

❋ Mezcladas: $f(\mathbf{X}) = \sum_{i=1}^{G} \pi_i f_i(\mathbf{X})$

❋ T^2 de Hotelling: si **X** es un vector con $N_p(\mu, \mathbf{V})$

$$(\mathbf{X}-\boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{X}-\boldsymbol{\mu}) \approx \chi^2_p;\ (\mathbf{X}-\boldsymbol{\mu})'\hat{\mathbf{S}}^{-1}(\mathbf{X}-\boldsymbol{\mu}) \approx T^2_{(p,n-1)}$$

$$F_{(p,n\text{-}p)} = \frac{n-p}{p(n-1)} T^2_{(p,n-1)}$$

ESTIMACIÓN MV

Este método escoge como estimador de los parámetros aquel valor que hace máxima la probabilidad de que el modelo a estimar genere la muestra observada.

Los estimadores MV son centrados, eficientes, asintóticamente normales, suficientes e invariantes.

Para v. normales p-dim., los estimadores MV de μ y **V** son $\overline{\mathbf{x}}$ y **S** (**S** es sesgado pero n**S**/(n-1) no lo es)

$$\overline{\mathbf{x}} \approx N_p(\boldsymbol{\mu}, \mathbf{V}/n) \qquad n\mathbf{S} \approx W_p(n-1, \mathbf{V})$$

RAZÓN DE VEROSIMILITUDES

Método para realizar contrastes sobre parámetros vectoriales, con $H_0{:}\theta\in\Omega_0$ frente a $H_1{:}\theta\in\Omega-\Omega_0$ y tomando el valor que hace más probable obtener la muestra observada y que es compatible con H_0.

Si $f(H_i)$ es el máximo valor de verosimilitud compatible con H_i $(i=0,1) \Rightarrow RV= f(H_0)/ f(H_1) \leq 1$.

Rechazamos H_0 cuando RV sea suficientemente pequeña. Cuando n es lo bastante grande este contraste equivale a $\lambda= D(H_0)-D(H_1) \approx\chi^2$, con D= desviación.

CONTRASTES PARA NORMAL

Sobre la media: $H_0:\mu=\mu_0$ (V cualq); $H_1: \mu\neq\mu_0$ (V cualq)

$$\lambda= n\ln\,[|S_0|/|S|] \approx\chi^2_p$$

Sobre la varianza:

Valor: $H_0:V=V_0$ (μ cualq); H_1: V y μ cualq $\Rightarrow \lambda= \ln\,[|V_0|/|S|]+n\,tr(V_0^{-1}S)-np \approx\chi^2_{p(p+1)/2}$

sfericidad: $H_0:V=\sigma^2 I$ (μ cualq); H_1: V y μ cualq $\Rightarrow \lambda= np\ln[\hat{\sigma}^2]+n\ln|S|-np \approx\chi^2_{[p(p+1)/2]-1}$

Independencia: H_0:V=diagonal (μ cualq); H_1: V y μ cualq $\Rightarrow\lambda= -n\ln|R| \approx\chi^2_{p(p+1)/2}$

OTROS CONTRASTES

De igualdad de medias:

H_0:μ_1=...=μ_G (V cualq); H_1: no todas iguales (V cualq para ambas)

$$\lambda = m\ln\left[|S|/|S_W|\right] \approx \chi^2_{p(G-1)}$$

con S_W estimación de la varianza común a los grupos

De datos atípicos (caso particular del anterior):

H_0: los datos todos de la misma poblac.; H_1: no todos lo son

Eliminar observaciones con $D^2 > \chi^2_{p;\,0{,}05}$

CONTRASTES DE NORMALIDAD

Univariante:

Si la variable es normal

$$X^2 = \frac{nA^2}{6} + \frac{n(K-3)^2}{24} \qquad > \chi^2_{2;\alpha}$$

con A coef. De asimetría y K de curtosis

Multivariante: definimos A_p y K_p planteando un contraste de baja potencia basado en ellos.

1. Se rechaza la hipótesis si existen datos atípicos
2. Se rechaza la hipótesis si las relaciones no son lineales

ANÁLISIS FACTORIAL

Spearman, psicólogo, propuso el primer modelo factorial para explicar la inteligencia mediante diferentes variables.

Charles Spearman

1863-1945

http://www.cps.nova.edu/~cpphelp/pantheon/spearman.htm

Tiene por objeto explicar un conjunto de variables observadas por medio de un pequeño número de variables latentes no observadas en la naturaleza que llamamos factores

*p.e. Utilizamos diversas pruebas de capacidad mental podemos preguntarnos si existen factores no directamente observables que expliquen los resultados. El conjunto de estos factores será lo que llamamos **inteligencia** y es importante conocer cuantas dimensiones tiene, cómo caracterizarlas y medirlas.*

ORIGEN DEL ANÁLISIS FACTORIAL

Se busca estimar la inteligencia de un grupo de 24 niños de una escuela de un pequeño pueblo.

Las medidas para estimar la inteligencia resultaron malas ya que empleaban variables las notas fuertemente correlacionadas dos a dos (p.e. en exámenes de matemáticas y letras).

Las correlaciones observadas entre las variables son todas positivas y siguen un patrón.

Sperman buscó un modelo que reflejara dicho patrón.

ORIGEN DEL ANÁLISIS FACTORIAL

Matriz de correlaciones entre notas de diferentes exámenes.

	Classics	French	English	Math	Discr	Music
Classics	1	.83	.78	.70	.66	.63
French	.83	1	.67	.67	.65	.57
English	.78	.67	1	.64	.54	.51
Math	.70	.67	.64	1	.45	.51
Discr	.66	.65	.54	.45	1	.40
Music	.63	.57	.51	.51	.40	1

ORIGEN DEL ANÁLISIS FACTORIAL

Una explicación a la correlación:

El modelo de puntuaciones de los test puede considerarse con una parte común a todas las pruebas y una parte específica de cada prueba.

Con las hipótesis correctas se puede obtener un modelo de correlaciones como el anterior.

El modelo de análisis factorial refleja esta idea básica.

LOGROS DEL ANÁLISIS FACTORIAL

- Una explicación a la correlación:
- Obtener unas nuevas variables, no correlacionadas que explican la estructura original de correlaciones.
- Permite una interpretación contextual de las nuevas variables
- Evalúa los datos originales a la luz de las nuevas variables.

NO HA CONSEGUIDO

- Muchos estadísticos no emplean este modelo factorial.
- Desgraciadamente las soluciones no son únicas
  - Hay que tomar muchas decisiones de forma subjetiva.

•Interpretación vectorial del A. Factorial

EL Análisis Factorial es útil para realizar una representación geométrica vectorial de las variables aleatorias.

Un vector viene dado por:

-módulo

-dirección

-sentido

- **Módulo**

Si representamos una variable mediante un vector, su módulo se puede representar como:

$$\|x_i\| = \sigma(x_i)$$

Ya que el módulo de un vector es $\|x_i\| = \sqrt{x_{i1}^2 + x_{i2}^2 + \ldots + x_{in}^2}$

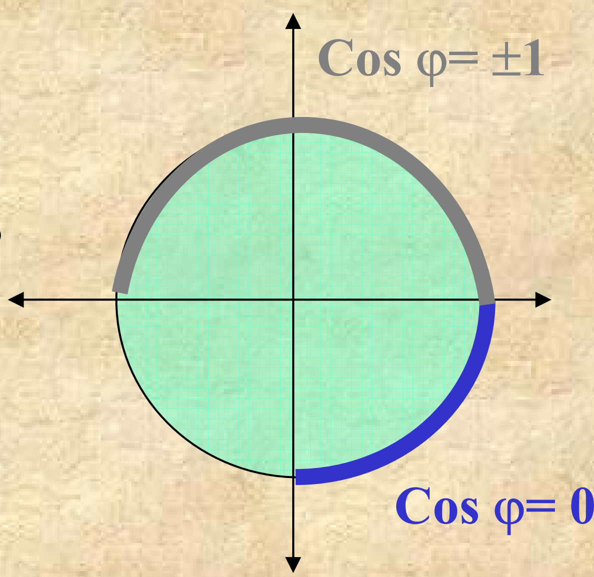
Y la desviación típica de la variable x=(x_{i1}, x_{i2},…,x_{in}) es:

$$\sigma(x_i) = \sqrt{\sum \left(x_{ij} - \bar{x}_i\right)^2} = \sqrt{\sum x_{ij}^2}$$

$$\bar{x}_i = 0$$

- **Dirección y Sentido**

- La dirección es una propiedad relativa a otro vector, que se supone fijo. Y se puede caracterizar mediante el ángulo φ que forman los dos vectores.
- Cosφ varía entre ±1.
- Si cos φ =±1, los dos vectores tienen la misma dirección.
- Si cos φ =0, los dos vectores son perpendiculares



- **Dirección y Sentido**

- Por otro lado dadas dos variables, x e y, el coeficiente de correlación ρ, verifica

  $-1 \leq \rho \leq 1$.

- Cuando $\rho=\pm 1$, entonces $y=ax$, los vectores tendrán la misma dirección. Y la relación lineal será mínima, es decir los vectores son perpendiculares cuando $\rho=0$.

- Por tanto se puede poner $\rho = \cos\varphi$

- **Dirección y Sentido**

- La correlación da una idea de la separación de dos variables como vectores.

	Misma dirección	Máxima separación
Variables	$\rho=\pm1$	$\rho=0$
Vectores	$\varphi=0º, 180º$ $\text{Cos}\varphi=\pm1$	$\varphi=90º$ $\text{Cos}\varphi=0$

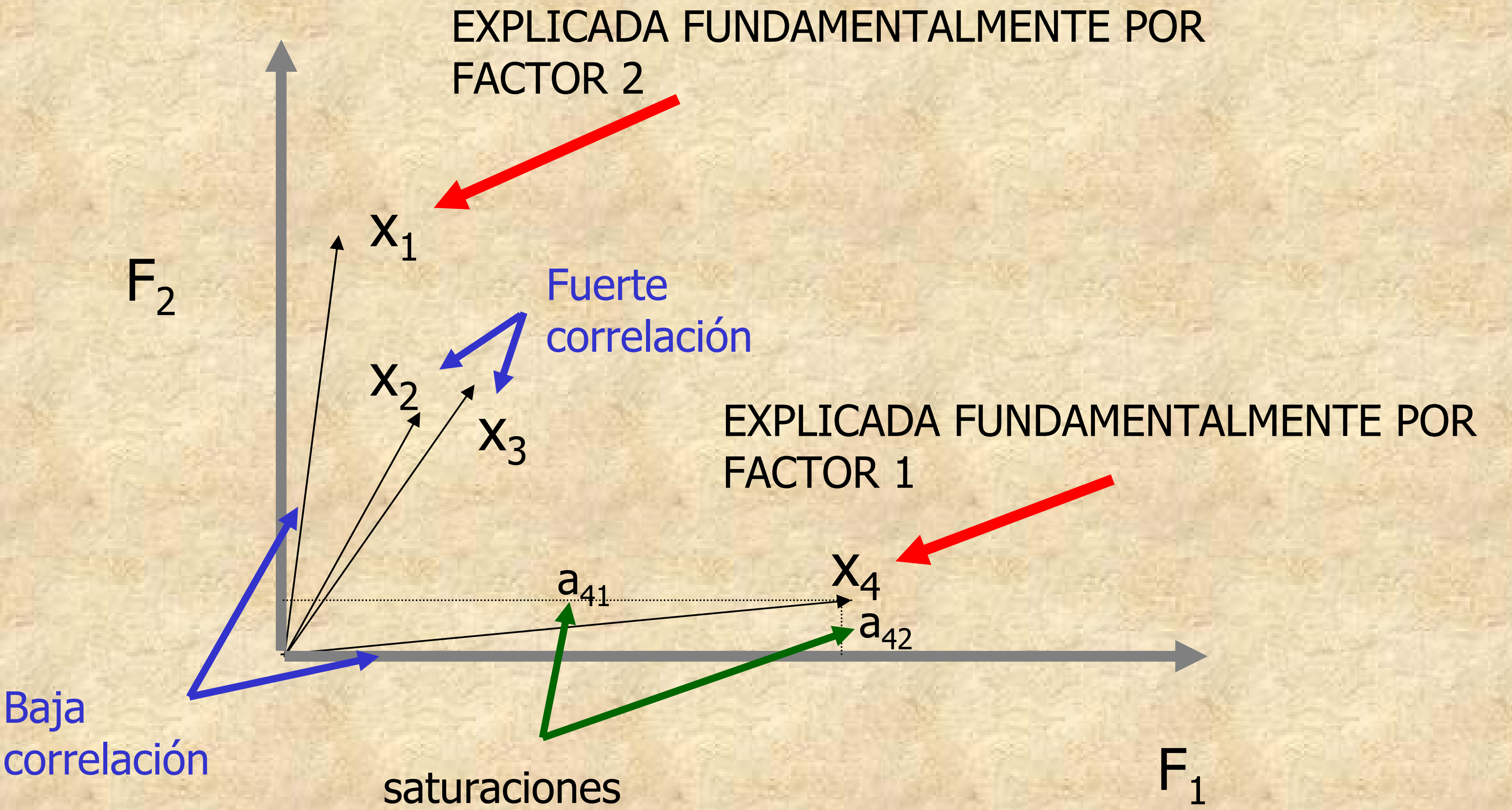
- **Dirección y Sentido**

- Si las variables son reducidas, estarán representadas por vectores de módulo unidad. Si las n variables son linealmente independientes (ρ=0) ocuparán un espacio n-dimensional.
- El análisis factorial pretende expresar n variables sin sus **unicidades**:

$$x_i^{'} = x_i - d_i u_i = a_{i1} F_1 + \ldots + a_{im} F_m$$

- $Var(x_i)=1=h_i^2+d_i^2$, si se elimina la unicidad el módulo es menor que 1.

• Dirección y Sentido



Fases de su aplicación

1º PREPARACIÓN DE LOS DATOS PARA EL ANÁLISIS

- Comprobación de los supuestos básicos
- Elección de las variables predictoras
- Elección de la matriz de datos:covarianzas y correlaciones

2º EXTRACCIÓN DE LOS FACTORES

- Procedimientos de extracción factorial
- Criterios de selección del número de factores

3º OBTENCIÓN DE LA MATRIZ E INTERPRETACIÓN

- La rotación de factores:ortogonal u oblicua

4º VALIDACIÓN DEL MODELO

- NEGATIVA → Introducción de modificaciones → REPLANTEAR
- POSITIVA → Conclusión del análisis
- POSITIVA → Cálculo de puntuaciones factoriales → Conclusión del análisis

R
E
P
L
A
N
T
E
A
R

EL MODELO FACTORIAL

- Hipótesis básicas:

Partimos de observaciones del vector $\mathbf{X}_{px1} \sim N(\mu, V)$

$$\mathbf{X} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f} + \mathbf{u}$$

donde - $\mathbf{f}_{\mathbf{mx1}}$ vector de factores no observados y sigue una $N_m(0,I)$

- Λ_{pxm} *matriz de carga* de constantes desconocidas con ($m<p$) coeficientes del modelo

- $\mathbf{u}_{px1}$ vector de perturbaciones que sigue una $N(0,\Psi)$, Ψ matriz diagonal independiente de f

EL MODELO FACTORIAL

Variables observadas

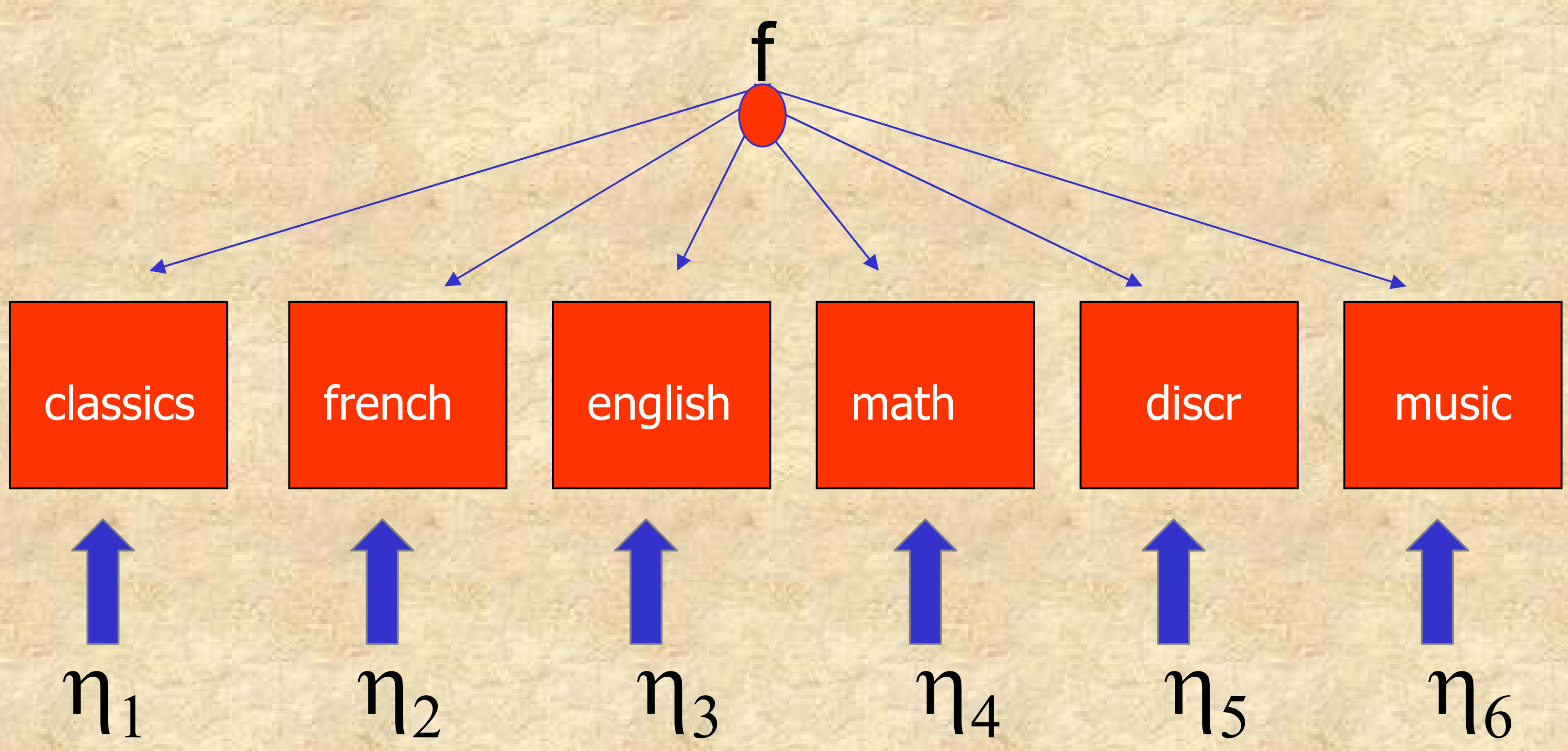
Factores comunes

$$x_i = \mu_i + \sum_{j=1}^{m} \lambda_{ij} f_j + \eta_i$$

Factores específicos

Factores de comunalidad

Esquemáticamente



EL MODELO FACTORIAL

- Propiedades:

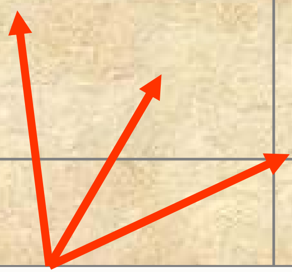
❶ La matriz de carga Λ contiene las covarianzas entre los factores y las variables observadas x. Si están estandarizadas, también contiene las correlaciones.

❷ La matriz **V** de las covarianzas entre las x se descompone en parte explicada por los factores y ruido: $\mathbf{V} = \Lambda \Lambda' + \Psi$

(*variabilidad observada= var. común+ var. específica*)

EL MODELO FACTORIAL

$\lambda^2_{11} + \psi_{11}$	$\lambda_{11}\ \lambda_{21}$	$\lambda_{11}\ \lambda_{31}$	$\lambda_{11}\ \lambda_{41}$	$\lambda_{11}\ \lambda_{51}$	$\lambda_{11}\ \lambda_{61}$
	$\lambda^2_{21} + \psi_{21}$	$\lambda_{21}\ \lambda_{31}$	$\lambda_{21}\ \lambda_{41}$	$\lambda_{21}\ \lambda_{51}$	$\lambda_{21}\ \lambda_{61}$
		$\lambda^2_{31} + \psi_{31}$	$\lambda_{31}\ \lambda_{41}$	$\lambda_{31}\ \lambda_{51}$	$\lambda_{31}\ \lambda_{61}$
			$\lambda^2_{41} + \psi_{41}$	$\lambda_{41}\ \lambda_{51}$	$\lambda_{41}\ \lambda_{61}$
				$\lambda^2_{51} + \psi_{51}$	$\lambda_{51}\lambda_{61}$
					$\lambda^2_{61} + \psi_{61}$



Can probably see a LOT of ways Spearman's pattern could emerge.

EL MODELO FACTORIAL

$var(x_1)$	$cov(x_1,x_2)$	$cov(x_1,x_3)$	$cov(x_1,x_4)$	$cov(x_1,x_5)$	$cov(x_1,x_6)$
	$var(x_2)$	$cov(x_2,x_3)$	$cov(x_2,x_4)$	$cov(x_2,x_5)$	$cov(x_2,x_6)$
		$var(x_3)$	$cov(x_3,x_4)$	$cov(x_3,x_5)$	$cov(x_3,x_6)$
			$var(x_4)$	$cov(x_4,x_5)$	$cov(x_4,x_6)$
				$var(x_5)$	$cov(x_5,x_6)$
					$var(x_6)$

=

$\lambda^2_{11} + \psi_{11}$	$\lambda_{11}\,\lambda_{21}$	$\lambda_{11}\,\lambda_{31}$	$\lambda_{11}\,\lambda_{41}$	$\lambda_{11}\,\lambda_{51}$	$\lambda_{11}\,\lambda_{61}$
	$\lambda^2_{21} + \psi_{21}$	$\lambda_{21}\,\lambda_{31}$	$\lambda_{21}\,\lambda_{41}$	$\lambda_{21}\,\lambda_{51}$	$\lambda_{21}\,\lambda_{61}$
		$\lambda^2_{31} + \psi_{31}$	$\lambda_{31}\,\lambda_{41}$	$\lambda_{31}\,\lambda_{51}$	$\lambda_{31}\,\lambda_{61}$
			$\lambda^2_{41} + \psi_{41}$	$\lambda_{41}\,\lambda_{51}$	$\lambda_{41}\,\lambda_{61}$
				$\lambda^2_{51} + \psi_{51}$	$\lambda_{51}\lambda_{61}$
					$\lambda^2_{61} + \psi_{61}$

Similar a la matriz de covarianzas

EL MODELO FACTORIAL

$1-\psi_{11}$	$corr(x_1,x_2)$	$corr(x_1,x_3)$	$corr(x_1,x_4)$	$corr(x_1,x_5)$	$corr(x_1,x_6)$
	$1-\psi_{21}$	$corr(x_2,x_3)$	$corr(x_2,x_4)$	$corr(x_2,x_5)$	$corr(x_2,x_6)$
		$1-\psi_{31}$	$corr(x_3,x_4)$	$corr(x_3,x_5)$	$corr(x_3,x_6)$
			$1-\psi_{41}$	$corr(x_4,x_5)$	$corr(x_4,x_6)$
				$1-\psi_{51}$	$corr(x_5,x_6)$
					$1-\psi_{61}$

=

λ^2_{11}	$\lambda_{11}\lambda_{21}$	$\lambda_{11}\lambda_{31}$	$\lambda_{11}\lambda_{41}$	$\lambda_{11}\lambda_{51}$	$\lambda_{11}\lambda_{61}$
	λ^2_{21}	$\lambda_{21}\lambda_{31}$	$\lambda_{21}\lambda_{41}$	$\lambda_{21}\lambda_{51}$	$\lambda_{21}\lambda_{61}$
		λ^2_{31}	$\lambda_{31}\lambda_{41}$	$\lambda_{31}\lambda_{51}$	$\lambda_{31}\lambda_{61}$
			λ^2_{41}	$\lambda_{41}\lambda_{51}$	$\lambda_{41}\lambda_{61}$
				λ^2_{51}	$\lambda_{51}\lambda_{61}$
					λ^2_{61}

Las comunalidades se obtienen con la correlación (datos estandarizados)

EL MODELO FACTORIAL

- Propiedades:

❸ El modelo está indeterminado frente a rotaciones.

❹ Para identificar Λ debemos poner restricciones

I. $\Lambda'\Lambda=\mathbf{D}$, matriz diagonal $\Rightarrow$ ortogonalidad entre f y x

II. $\Lambda'\,\Psi^{-1}\Lambda=\mathbf{D}$, diagonal $\Rightarrow$ los factores ponderados por las varianzas de las perturbaciones y x son incorrelados.

❺ Para poder resolver el sistema con las condiciones anteriores $(p-m)^2 \geq p+m$.

PREPARACIÓN DE DATOS

1º Número suficiente de datos (más de 200).

2º Comprobación de la **Normalidad** multivariante.

3º Relaciones lineales entre variables y **correlaciones** > 30%

4º La matriz de correlaciones a usar tiene como comunalidades los **coeficientes de determinación**.

Si utilizamos comunalidades 1, estamos usando ACP.

5º Analizar las matrices de **correlaciones parciales**:si las variables comparten factores comunes las correlaciones parciales deben ser próximas a 0.

SELECCIÓN DEL Nº DE FACTORES

1º Autovalores (o varianza total explicada por cada factor)

2º Porcentaje de varianza total atribuible a cada factor.

3º El gráfico de sedimentación.

4º Significatividad.

5º Interpretabilidad.

ESTIMACIÓN DEL MODELO

Se pueden emplear diferentes métodos de estimación:

1. Método del factor principal: se basa en los métodos de componentes principales, busca los autovectores.
2. Máxima verosimilitud: la estimación no siempre converge, es invariante frente a transformaciones lineales y no usa las varianzas.
3. Mediante mínimos cuadrados con el algoritmo EM.

ROTACIÓN DE FACTORES

Ya obtenida la matriz de factores:

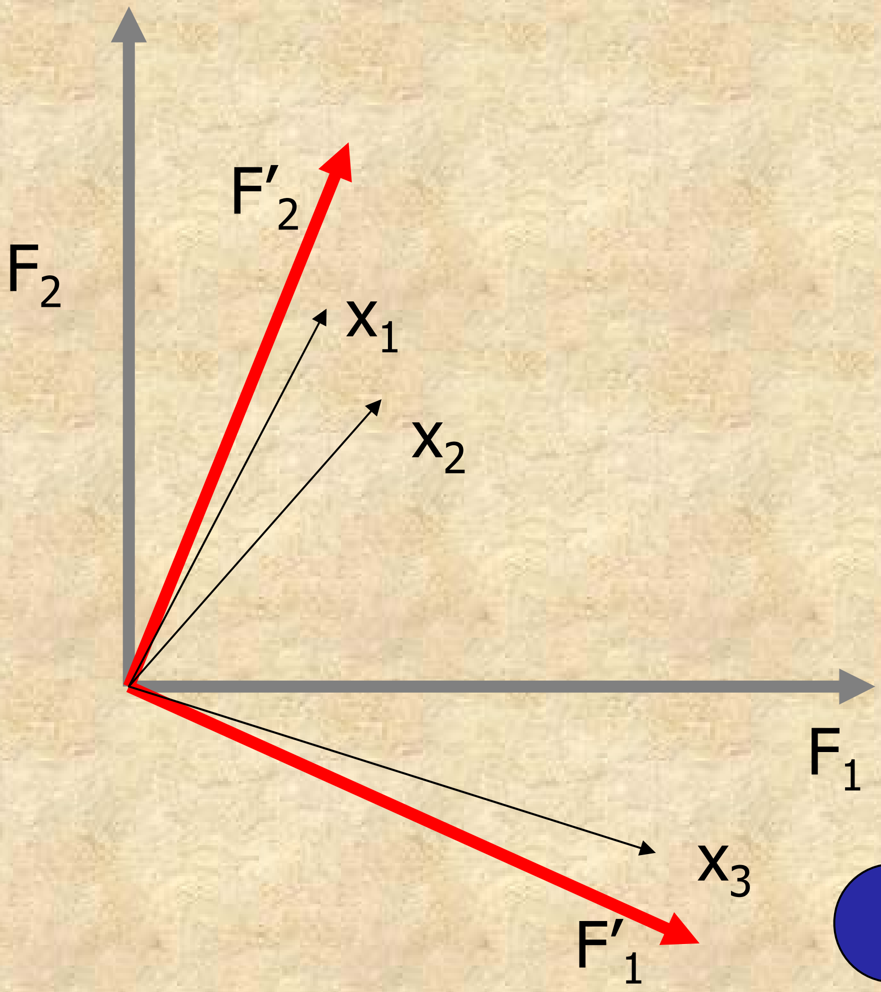
$$\begin{pmatrix} a_{11} \cdots a_{1m} \\ \cdots \\ a_{n1} \cdots a_{nm} \end{pmatrix}$$

Coeficiente factorial de la v. 1 en el factor m = $\lambda^2{}_{1m}$

El siguiente paso es la rotación de los factores para que sea más interpretable el resultado.

Se elegirá el número mínimo de factores compatible con las variables, y entre las diferentes rotaciones, aquellos cuya estructura sea más simple.

ROTACIÓN DE FACTORES



"Estructura simple"

Cada factor con pocos coef. factoriales elevados y los demás próximos a cero.

Cada variable un coef. factorial elevado sólo en un factor.

Distinta distribución de todos los factores.

ROTACIÓN DE FACTORES

- **Métodos para realizar la rotación:**
  - Factores ortogonales: Incorrelados y por tanto perpendiculares entre sí.
    - Quartimax. Suma de las cuartas potencias de las saturaciones máxima. $Q = \sum \sum a_{ij}^4$
    - **Varimax**. Suma de las varianzas de los cuadrados de las saturaciones máxima. $V = \sum \sum a_{ij}^2$
    - Equimax. Intermedia de las anteriores
  - Factores oblicuos: Están correlacionados entre sí. Se busca el c.d.g. de los grupos de variables, y se obtienen los nuevos factores con fuerte saturación.

VALIDACIÓN

- Contraste de verosimilitud:

H_0 : **V**= Λ Λ'+ Ψ , bilateral, se rechaza H_0 si

$$\left(n-1-\frac{2p+4m+5}{6} \right) \ln \frac{\left| \boldsymbol{\Lambda\Lambda}' - (\mathbf{S}-\hat{\boldsymbol{\Psi}}) \right|}{|\mathbf{S}|} > \chi^2_{\frac{(p-m)^2-(p+m)}{2}}$$

- Diagnosis del modelo:

Mediante los residuos: deben ser Normales, de media 0 y su matriz de covarianzas debe ser diagonal.

- Medida del ajuste: $R^2 = 1 - \frac{\left|\widehat{\Psi}\right|^{1/p}}{\left|\widehat{V}\right|^{1/p}}$

El análisis factorial está relacionado con los componentes principales

DIFERENCIAS

1. Explica las correlaciones y las CP las varianzas

2. Los errores están incorrelados y en las CP no

3. Se construye un modelo

SIMILITUDES

1. Reducen la dimensión del problema

2. Permiten construir nuevas variables

3. Conducen a resultados similares si $\Psi \cong 0$

USO DEL A. FACTORIAL

- Exploratoria:

Para encontrar factores que expliquen las variables.

- Confirmatoria:

Para contrastar teorías, con nº de factores conocido, se contrasta la igualdad a cero de alguno o la igualdad entre sí.

Los contrastes son muy restrictivos y sensibles a la falta de normalidad.

ULTIMAS APLICACIONES

Estudios atmosféricos (año 2005) :

Investiga estructuras relativas al comportamiento humano y la climatología en la concentración de ozono, debida al tráfico, en el sur de California.

Ecología (año 2005) :

Para encontrar estructuras que expliquen la competencia por el pasto entre fígaros y moscas de Japón.

Ejemplo 1:

Se tienen 72 observaciones de distintas variables de calidad del aire en tres estaciones madrileñas.

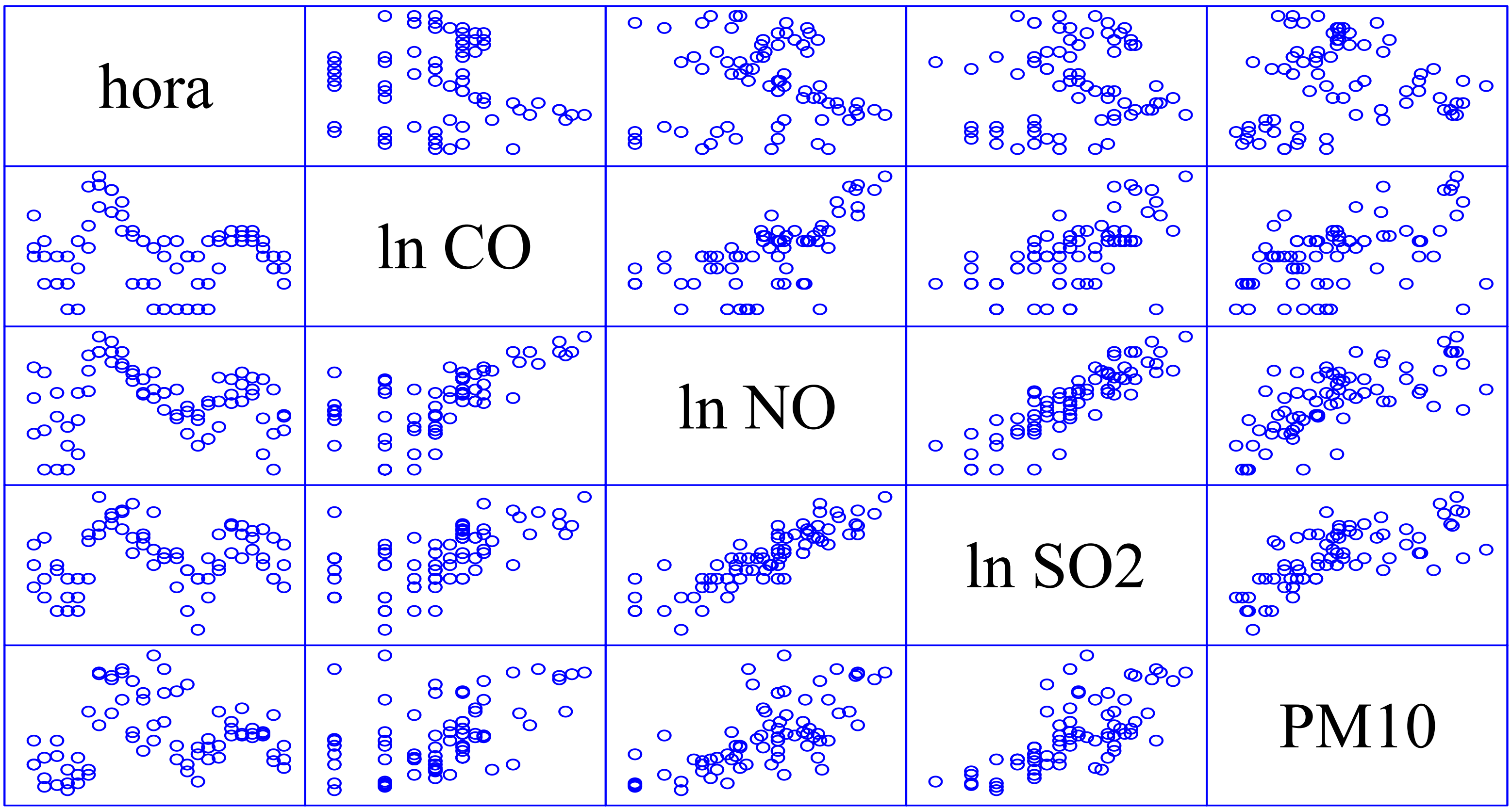
Las variables medidas son concentraciones, por metro cúbico de aire, de diferentes agentes contaminantes:

Cada hora durante un día (17-2-02) se miden CO, NO_2, NO_x, SO_2, y partículas suspendidas.

Se transforman las variables asimétricas y se descarta NO_x porque está muy correlacionada con NO_2.

Ejemplo 1:

El gráfico de las variables a considerar es:

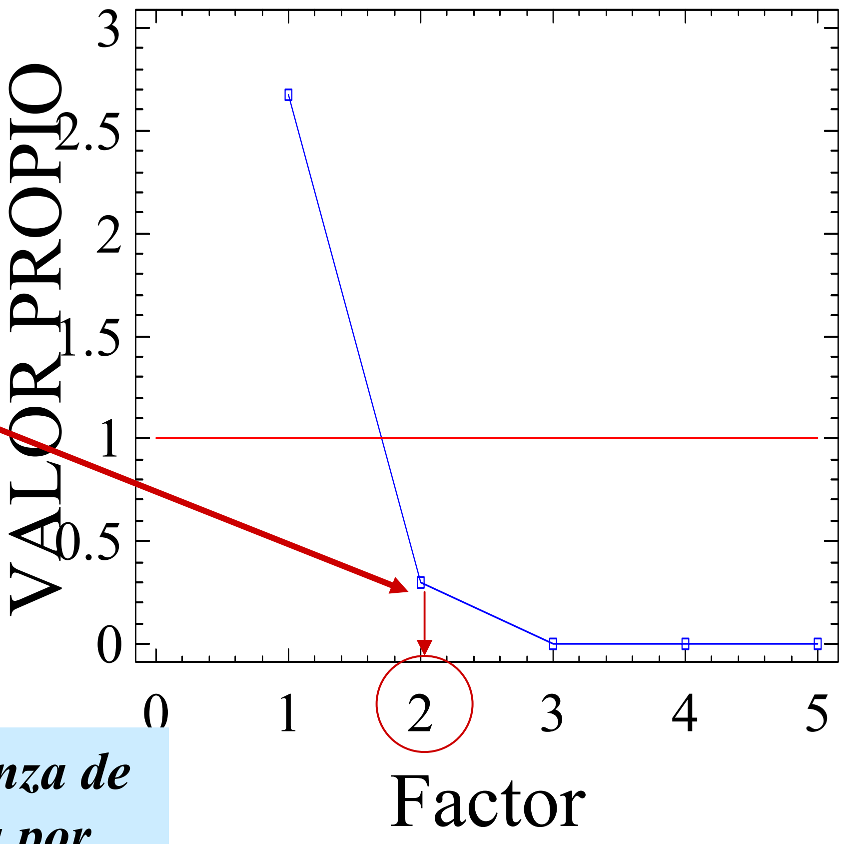


Ejemplo 1:

Factor Analysis

Factor Nº	Autovalor	%Varianza	%Acumulada
1	2,67081	89,952	89,952
2	0,29835	10,048	100,000
3	0,0	0,000	100,000
4	0,0	0,000	100,000
5	0,0	0,000	100,000

Variable	Comunalidad inicial
hora	0,161119
ln SO2	0,802419
ln CO	0,431071
ln NO	0,812834
PM10	0,514851

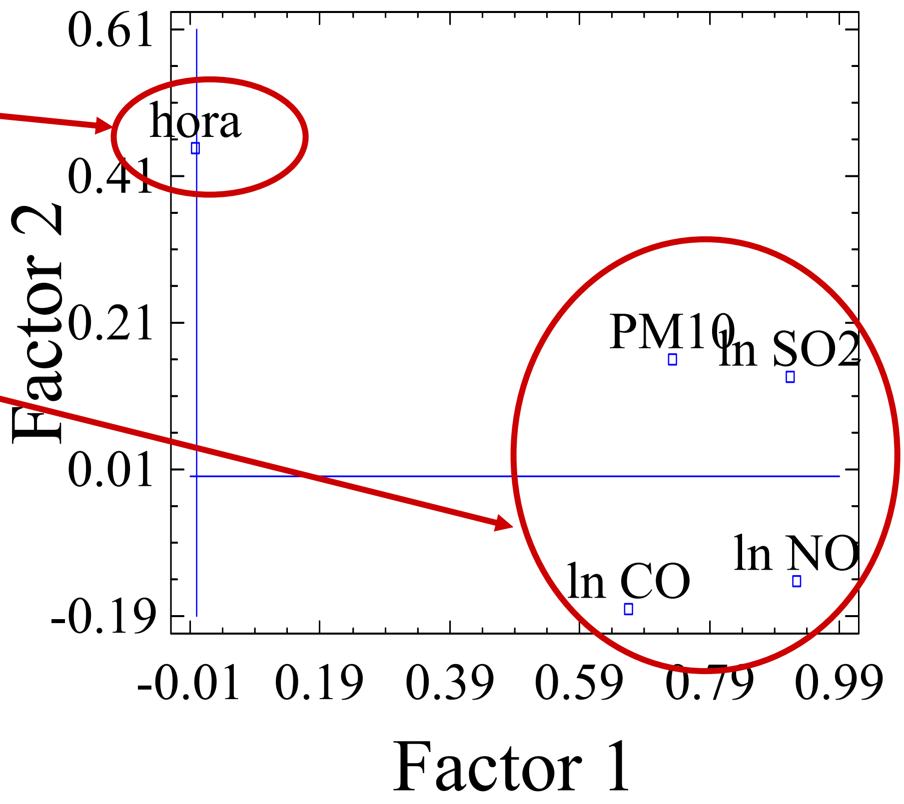


Porción de varianza de cada v. explicada por los factores comunes.

Ejemplo 1:

Variable	Factor 1	Factor 2
	------------	------------
hora	-0,0032736	-0,448764
ln SO2	0,914466	-0,140679
ln CO	0,665746	0,178111
ln NO	0,925136	0,137886
PM10	0,731748	-0,162573

Variable	Comunalidad estimada
hora	0,2014
ln SO2	0,856038
ln CO	0,474941
ln NO	0,87489
PM10	0,561886



Ejemplo 1:

Después de rotar las variables (VARIMAX):

Sin cambios.

Factor	Factor 1	Factor 2
hora	-0,0032736	-0,448764
ln SO2	0,914466	-0,140679
ln CO	0,665746	0,178111
ln NO	0,925136	0,137886
PM10	0,731748	-0,162573

Factor	Factor 1	Factor 2
hora	-0,00164928	0,448773
ln SO2	0,914969	0,137369
ln CO	0,665096	-0,180519
ln NO	0,924631	-0,141234
PM10	0,732332	0,159924

Ejemplo 2:

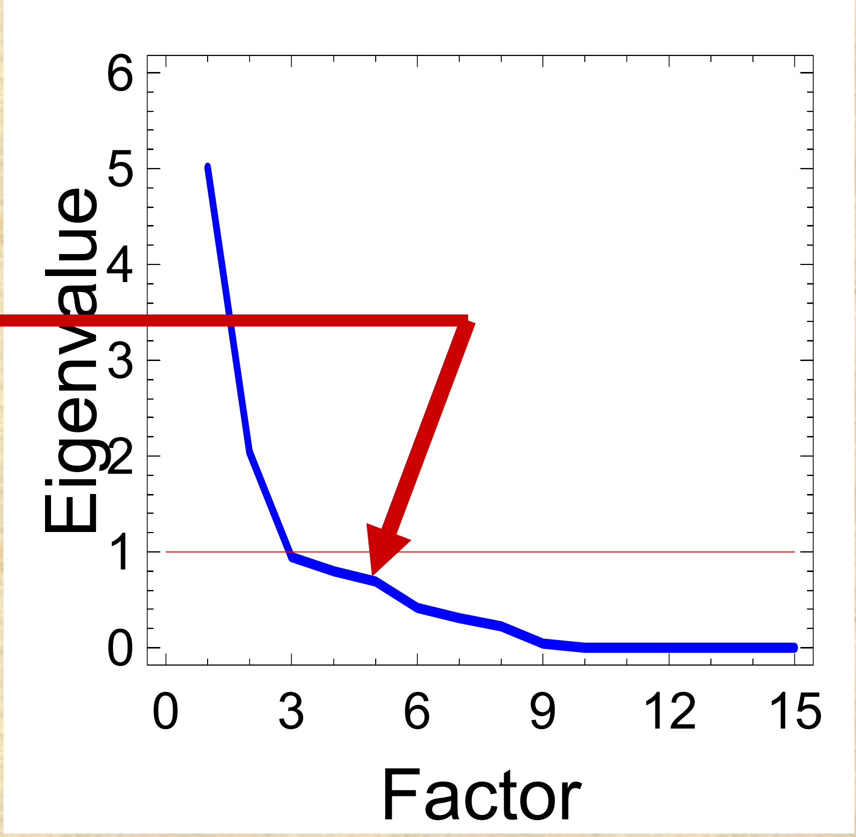
- Gunst y Mason (1980) toman 15 medidas antropométricas y de rendimiento físico a 50 policías de una gran ciudad.

- Las correlaciones lineales son altas (>0,3) salvo con la reacción visual y presión diastólica. Las corr. parciales son todas <0,4

Ejemplo 2:

Factor Analysis

Factor	autovalor	%V	%V acum.
1	5,01278	47,871	47,871
2	2,04749	19,553	67,424
3	0,946747	9,041	76,465
4	0,793066	7,574	84,038
5	0,690385	6,593	90,631
6	0,416064	3,973	94,605
7	0,304284	2,906	97,511
8	0,219814	2,099	99,610
9	0,040874	0,390	100,000
10	0,0	0,000	100,000
11	0,0	0,000	100,000
12	0,0	0,000	100,000
13	0,0	0,000	100,000
14	0,0	0,000	100,000
15	0,0	0,000	100,000



Ejemplo 2:

- Se decide utilizar todos los parámetros
- El número de factores elegido es 5
- Elegimos la estimación de máxima verosimilitud (método clásico)
- El método de rotación “varimax”

Ejemplo 2:

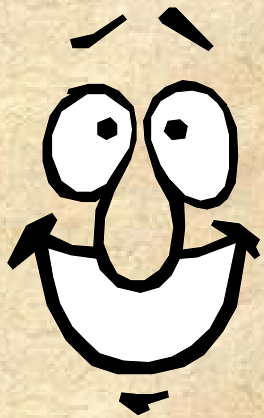
Variable	Communality
reaccion visual	0,433273
estatura	0,746427
peso	0,940737
ancho espalda	0,739887
a pelvis	0,728866
torax	0,892508
espesor piel	0,842592
pulso	0,472639
pres diastólica	0,358759
nº flexiones	0,715872
capacidad pulmonar	0,506498
frec pulso	0,583405
velocidad	0,622358
aguante	0,286925
grasa cuerpo	0,937507

Las comunalidades son altas por lo que el modelo obtenido puede ser adecuado

Ejemplo 2:

Matriz factorial original

variables	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
reaccion visual	0,104753	0,184032	0,0683623	-0,332853	0,544737
estatura	0,672771	-0,360882	0,349974	-0,24831	0,156148
peso	0,961668	-0,0648953	-0,0386423	0,130012	0,00904715
ancho espalda	0,662464	-0,349141	0,304915	0,0819886	-0,188232
a pelvis	0,647828	-0,32159	0,159467	0,3399	0,188134
torax	0,826678	-0,0097753	-0,0781708	0,327624	-0,0539064
espesor piel	0,646615	0,496019	-0,245036	-0,179969	-0,107692
pulso	-0,24812	0,480605	0,423153	-0,0489768	-0,00535452
pres diastólica	-0,0683411	0,323128	0,108775	0,373766	0,34555
nº flexiones	-0,645652	-0,361567	0,221189	0,210725	0,00697844
capacidad pulmonar	0,527637	-0,069253	0,337881	-0,285158	0,0189415
frec pulso	-0,052274	0,573175	0,444149	0,0383129	-0,339814
velocidad	-0,068812	-0,683927	-0,132073	-0,236085	-0,151017
aguante	-0,41048	-0,060454	-0,042545	0,128251	0,160414
grasa cuerpo	0,852566	0,369949	-0,251083	-0,0161583	0,0008112



Ejemplo 2:

Matriz factorial rotada

variables	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
reaccion visual	0,104248	0,014105	0,00274885	0,12171	0,656625
estatura	0,144587	0,787356	-0,102413	-0,223664	0,300052
peso	0,636728	0,696878	-0,213779	0,0856838	-0,0582384
ancho espalda	0,166665	0,781542	-0,0463495	-0,160487	-0,171449
a pelvis	0,137029	0,733972	-0,282289	0,240312	-0,0672869
torax	0,567733	0,583435	-0,18952	0,21767	-0,232015
espesor piel	0,861718	0,078856	0,1146	0,00244919	0,0791144
pulso	-0,103912	-0,139151	0,629782	0,160411	0,146672
pres diastólica	-0,0394324	-0,0473322	0,126544	0,594155	0,0848031
nº flexiones	-0,770092	-0,147179	-0,0162915	0,0182477	-0,160231
capacidad pulmonar	0,225718	0,547348	0,142048	-0,212415	0,251393
frec pulso	0,110744	-0,0317572	0,778385	0,0843323	-0,138577
velocidad	-0,284018	0,115854	-0,431487	-0,530976	-0,0791911
aguante	-0,344756	-0,259727	-0,0805441	0,152155	0,0136573
grasa cuerpo	0,901933	0,300465	-0,0514621	0,135391	0,0478926

Ejemplo 2:

¿Qué se observa en los factores?

Factor I	Peso, torax, espesor de piel, grasa del cuerpo y negativamente flexiones
Factor II	Estatura, peso, capacidad pulmonar, torax, anchura de espalda y pelvis
Factor III	Pulso y frecuencia de pulso
Factor IV	Presión diastólica y negativamente la velocidad
Factor V	reaccion visual

Ejemplo 2:

¿Qué se observa en los factores?

Talla	**Peso, torax, espesor de piel, grasa del cuerpo y negativamente flexiones**
Envergadura	**Estatura, peso, capacidad pulmonar, torax, anchura de espalda y pelvis**
Actividad	**Pulso y frecuencia de pulso**
Fuerza	**Presión diastólica y negativamente la velocidad**
Factor trivial	**reaccion visual**

Eliminar y repetir

Ejemplo 2:

Matriz factorial original con CP

variables	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
reaccion visual	0,115771	0,23649	-0,127622	-0,0616796	-0,900823
estatura	0,697829	-0,337249	-0,419024	-0,0597184	-0,228197
peso	0,951866	-0,0444574	0,0767846	-0,100931	0,0610142
ancho espalda	0,685652	-0,32752	-0,31506	-0,11806	0,231965
a pelvis	0,671231	-0,299374	-0,0879763	-0,487107	0,109135
torax	0,824163	0,005584	0,146635	-0,159021	0,188939
espesor piel	0,649055	0,475925	0,263523	0,227961	-0,0131707
pulso	-0,272585	0,591604	-0,529906	-0,0202328	0,00285753
pres diastólica	-0,0817311	0,429164	0,0490836	-0,774715	-0,0528924
nº flexiones	-0,666793	-0,365783	-0,274776	-0,240072	0,125651
capacidad pulmonar	0,576319	-0,0503616	-0,503535	0,171492	-0,177682
frec pulso	-0,058222	0,657629	-0,458096	0,114922	0,417287
velocidad	-0,0670392	-0,762356	-0,0313468	0,219866	-0,0445652
aguante	-0,46683	-0,083117	0,149595	-0,36143	-0,0919711
grasa cuerpo	0,841053	0,347341	0,267382	0,0723551	-0,0154591

Ejemplo 2:

Talla	**Estatura, peso, capacidad pulmonar, torax, anchura de espalda y pelvis, espesor de piel, grasa del cuerpo y negativamente flexiones**
Actividad	**Pulso, frecuencia de pulso y negativamente la velocidad**
Resistencia	**Pulso y capacidad pulmonar**
Fuerza	**Presión diastólica**
Factor trivial	**reaccion visual**

Resultado distinto: diferentes factores, diferente interpretación