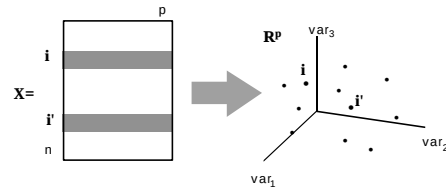
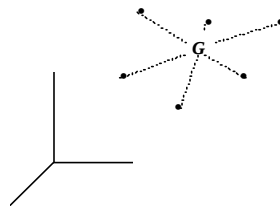


Análisis de Componentes Principales

- Nube de puntos



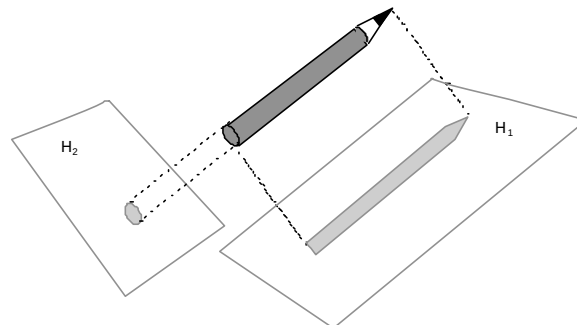
- Información de la nube de puntos: *La inercia*



1

- Objetivo:

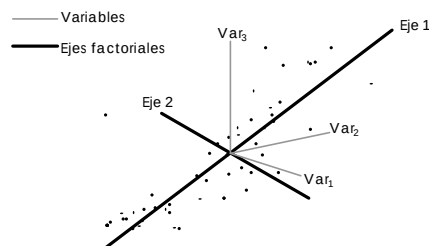
- Proyectar la nube de puntos sobre un subespacio (un plano) que conserve el máximo de información de la nube original.



2

Análisis de Componentes Principales

- Criterio matemático
 - Encontrar el subespacio que maximice la inercia proyectada.



- Descomposición de la inercia en direcciones ortogonales

$$I_{total} = I_1 + I_2 + \dots + I_p$$

$$I_1 > I_2 > \dots > I_p$$

3

Ajuste en R^p

Matriz de Datos X : **centrados** o **estandarizados**

Búsqueda de la dirección $u \in R^p$ maximizando la inercia: $y = X u$

$$\left. \begin{array}{l} \text{Max}_u \sum_{i=1}^n p y_i^2 = y' N y = u' X' N X u \\ u' u = 1 \end{array} \right\} X' N X u = l u$$

Diagonalización de la matriz de correlaciones (o covariancias)

$$\text{diag}(X' N X) = \begin{cases} \text{Cov}(X) \\ \text{Cor}(X) \end{cases} \rightarrow \begin{matrix} l_1, \dots, l_r \\ u_1, \dots, u_r \end{matrix} \quad r = \text{rang}(X)$$

$$\text{Max}_u u' X' N X u = l_1$$

4

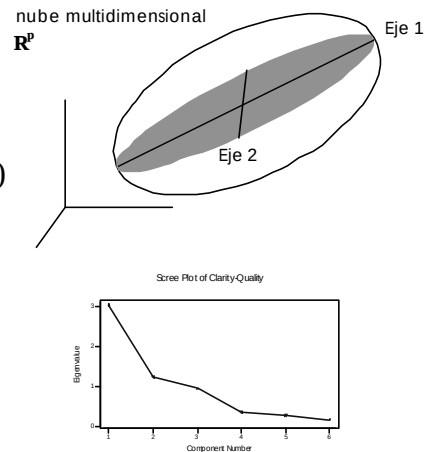
La dirección que maximiza la inercia proyectada es u_1 .
 La dirección que maximiza la inercia proyectada ortogonal a u_1
 es u_2 , ...

$$y_a = Xu_a$$

(componentes principales)

$$y_a' N y_a = I_a$$

Scree plot of eigenvalues:



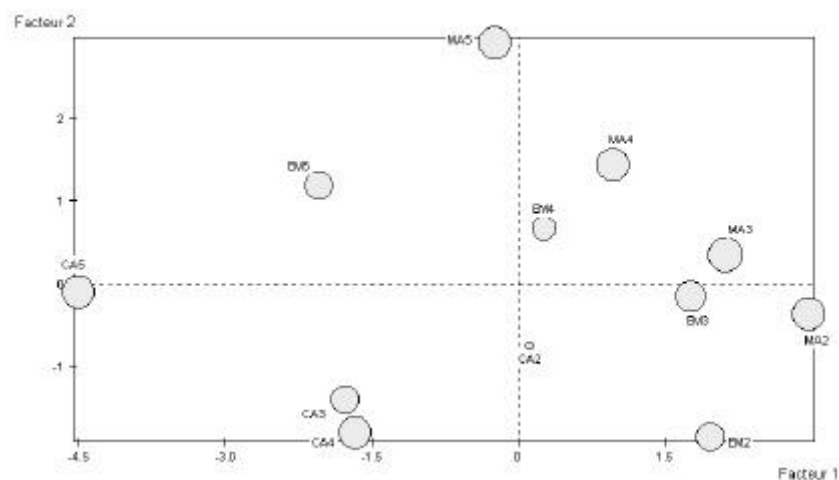
5

Datos: Consumo en alimentación por tipo de familia,
 según CAPS y número de hijos

Iden	Pan	Verdura	Fruta	Carne	Arroz	Leche	Vino	CAPS	N hijos
MA2	332	428	354	1437	526	247	427	1	1
EM2	293	559	388	1527	567	239	258	2	1
CA2	372	767	562	1948	927	235	433	3	1
MA3	406	563	341	1507	544	324	407	1	2
EM3	386	608	396	1501	558	319	363	2	2
CA3	438	843	689	2345	1148	243	341	3	2
MA4	534	660	367	1620	638	414	407	1	3
EM4	460	699	484	1856	762	400	416	2	3
CA4	385	789	621	2366	1149	304	282	3	3
MA5	655	776	423	1848	759	495	486	1	4
EM5	584	995	548	2056	893	518	319	2	4
CA5	515	1097	887	2630	1167	561	284	3	4

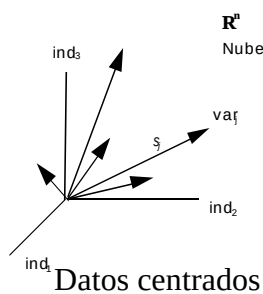
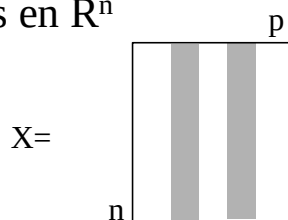
6

Primer plano factorial. Consumo en alimentación por tipo de familia, según CAPS y número de hijos

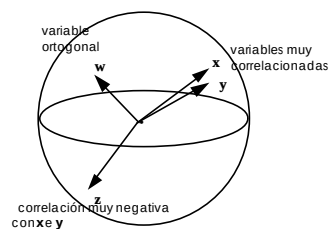


7

Nube de puntos en $\mathbb{R}^n$



$\mathbb{R}^n$
Nube de las variables

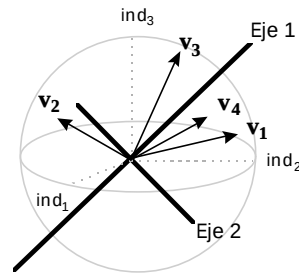


Datos centrados y reducidos

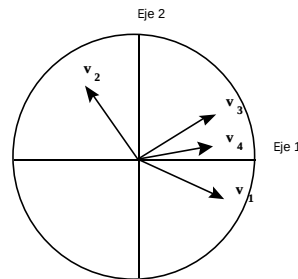
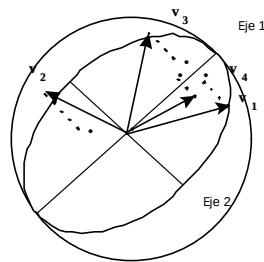
8

Análisis en R^n datos centrados y reducidos

Nube original



Primer plano factorial



Visualización óptima de las correlaciones entre variables

9

Ajuste en R^n

Matriz de Datos X : **centrados o estandarizados**

Búsqueda de la dirección $v \in R^p$ maximizando la inercia: $j = X'N^{-1/2}v$

$$\left. \begin{aligned} \text{Max}_v \sum_{j=1}^p j_j^2 = j'j = v'N^{-1/2}XX'N^{-1/2}v \\ v'v = 1 \end{aligned} \right\} N^{-1/2}XX'N^{-1/2}v = l v$$

Relaciones de transición entre ambos ajustes:

$$j_a = X'N^{-1/2}v_a$$

$$u_a = l^{-1/2}X'N^{-1/2}v_a$$

$$j_a'j_a = l_a$$

$$v_a = l^{-1/2}N^{-1/2}Xu_a$$

Fórmulas indirectas

Relación entre proyecciones

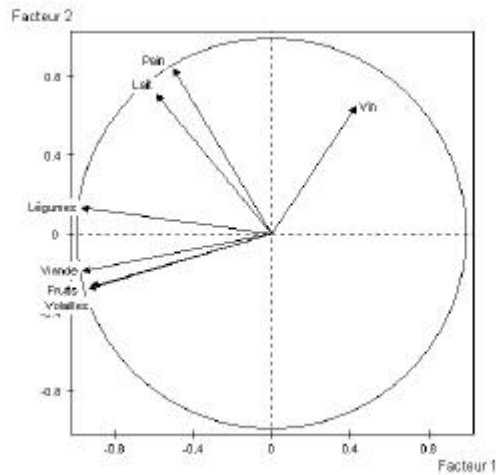
$$j_a = l^{1/2}u_a$$

$$j_a = l^{-1/2}X'Ny_a = \begin{cases} \text{cor}(x_j, y_a) \\ s_j \text{cor}(x_j, y_a) \end{cases}$$

$$y_a = l^{1/2}N^{-1/2}v_a$$

10

Primer plano factorial : Consumos en alimentación por tipo de familia



11

Representaciones simultáneas en ACP: *el biplot*

Descomposición en valores singulares en ACP

$$\text{eig}(X'NX) \equiv N^{1/2}X = V\Lambda^{1/2}U'$$

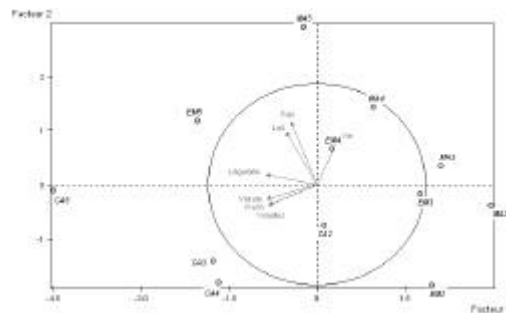
Biplot en R^p

$$X = N^{-1/2}V\Lambda^{1/2}U' = \Psi U'$$

Biplot en R^n

$$X = N^{-1/2}V\Phi' = \Psi_s \Phi'$$

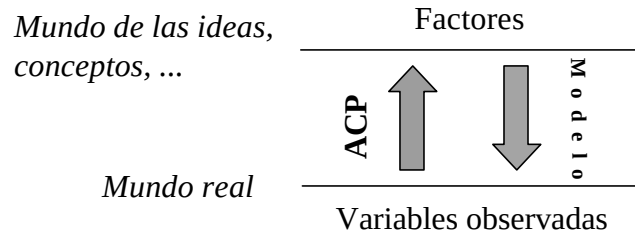
Ψ_s CP estandarizados



12

El ACP como modelo

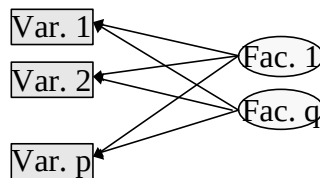
- El ACP es una técnica para ayudar a extraer los factores no observables, subyacentes en las variables observables.



$$\text{ACP: } \Psi_a = u_{1a} \mathbf{x}_1 + u_{2a} \mathbf{x}_2 + \dots + u_{pa} \mathbf{x}_p$$

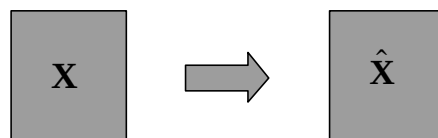
13

ACP como modelo



$$\hat{\mathbf{X}}_{(n,p)} = \Psi_{(n,q)} \mathbf{U}_{(q,p)}$$

$$\text{Modelo: } \mathbf{x}_j = u_{j1} \Psi_1 + u_{j2} \Psi_2 + \dots + u_{jq} \Psi_q + \mathbf{e}_j$$



Reconstitución óptima: $|\mathbf{X} - \hat{\mathbf{X}}|^2 = \sum_{a=q+1}^r l_a$

14