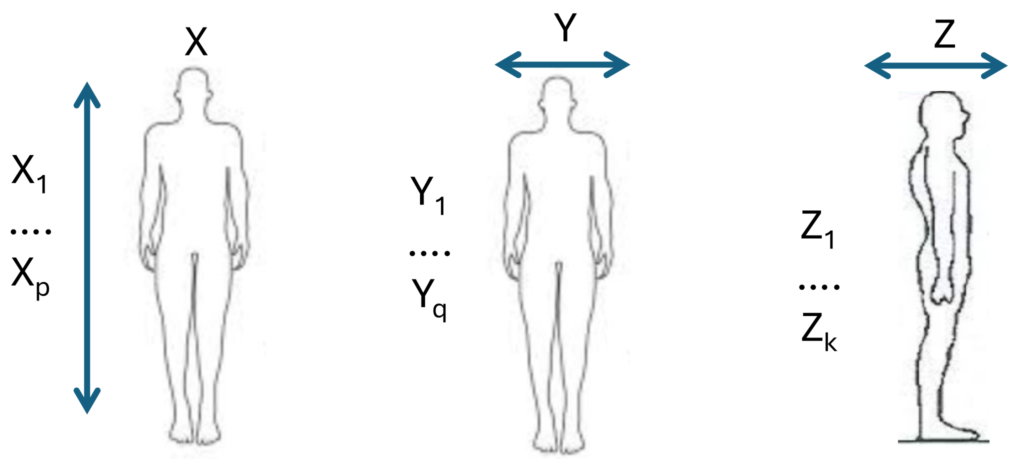
Componentes Principales

Interés: Reducir dimensiones (variables) en un problema

1

¿Qué es reducir dimensiones?



Dimensión 1 alto

Dimensión 2 ancho

Dimensión 3 fondo

2

¿Cómo se representan las dimensiones?

$X_1, ..X_p$ vectores de $\mathbb{R}^n$

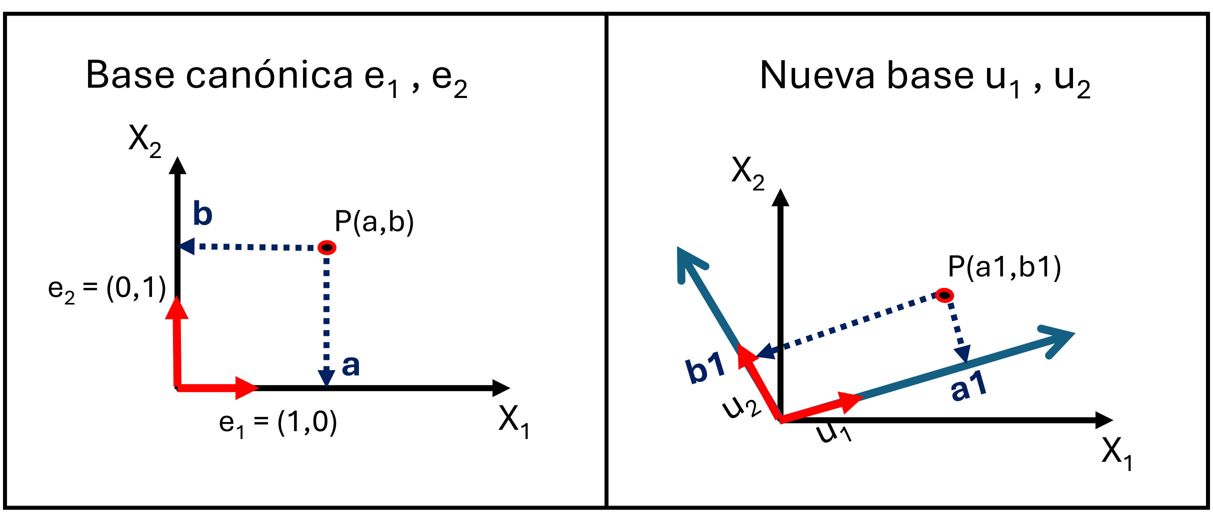
$\Re\{X_1, ..X_p\}$ sub espacio vectorial generado por la familia

Dimensión:
número máximo de vectores linealmente independientes en el espacio o sub espacio vectorial

Con la base se identifican los elementos del espacio mediante sus coordenadas relativas a una base

3

Caso p = 2 Proyecciones ortogonales) $P \in \Re\{X_1, ..X_p\}$

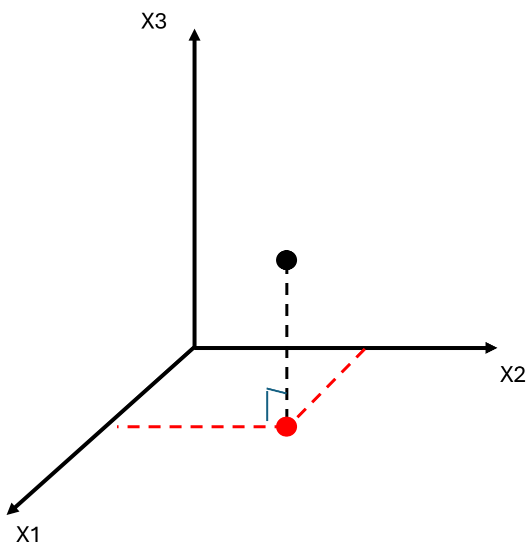


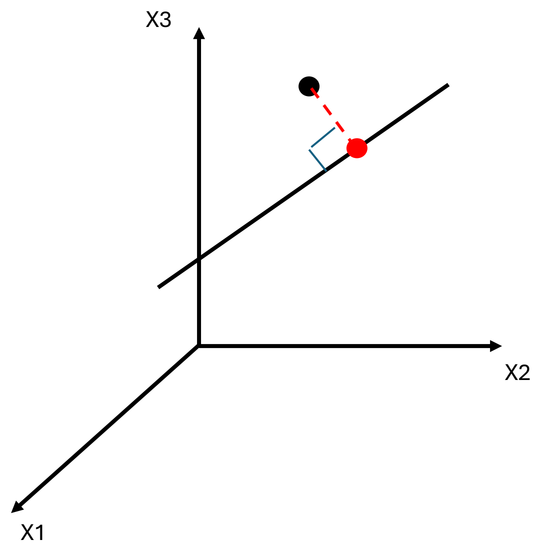
$$P = a\ e_1 + b\ e_2$$

$$P = a1\ u_1 + \text{b1}\ u_2$$

4

Entonces ¿Qué es una proyección (ortogonal)?





5

Los datos

Variables a analizar en el problema:

$$\mathbb{X} = (X_1, ..X_p)^t$$

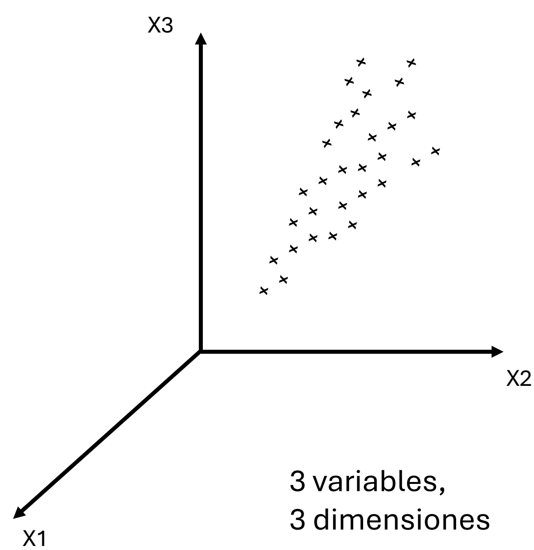
Matriz de datos:

$$X = (x_{ij})_{n,p} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$$

Muestra aleatoria de $\mathbb{X}$

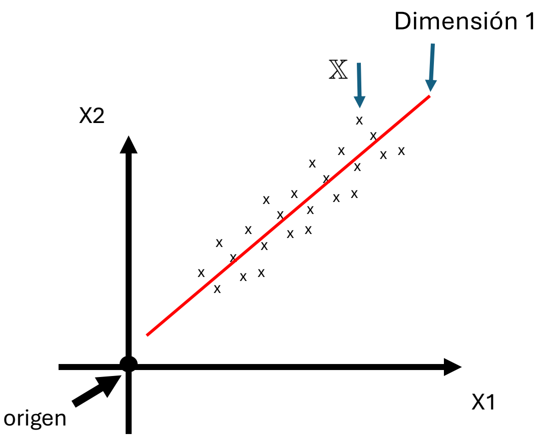
$$(X_i)_{i=\overline{1,n}}$$

Nube de puntos (datos)

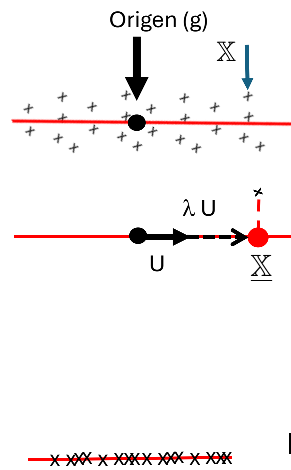


6

Idea:
Proyectar la nube de puntos en un subespacio vectorial centrado en una posición conveniente (espacio afín) con una dimensión menor.
Deformando la nube lo menos posible



Componentes principales

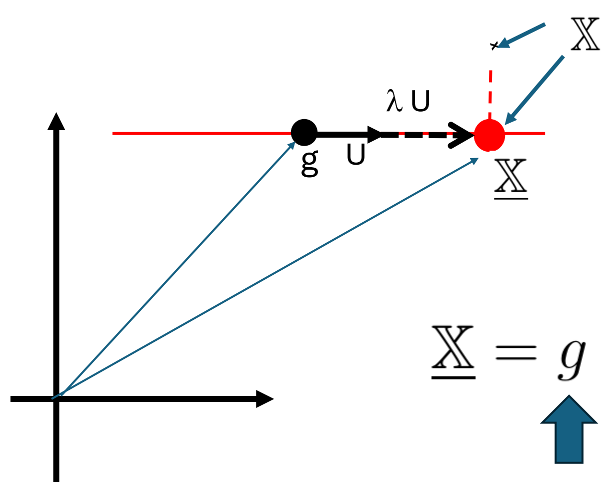


U base
λ U vector proyección
λ nuevas coordenadas

Nube de puntos en el espacio de dimensión reducida

7

Una propiedad importante de la proyección

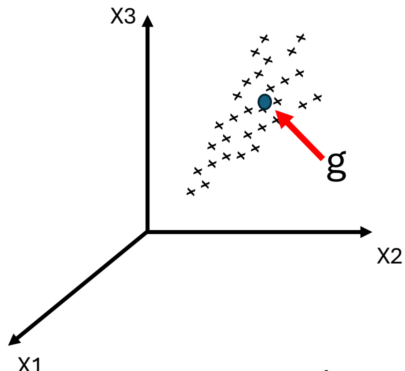


$$\underline{\mathbb{X}} = g + \lambda\, U$$

Recuerde esto

8

Idea:
Proyectar la nube de puntos en un subespacio vectorial centrado en una posición conveniente (espacio afín) con una dimensión menor.
Deformando la nube lo menos posible



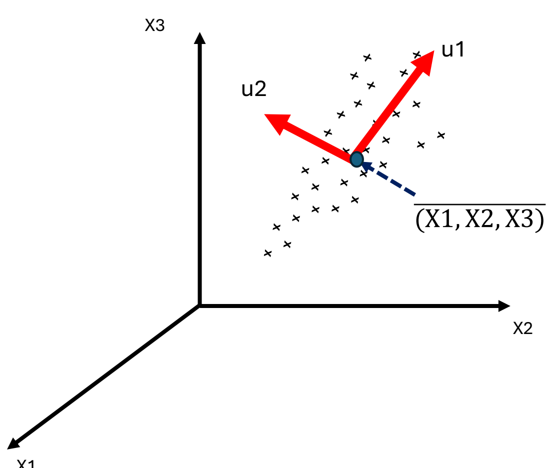
g centro

Para medir la forma (dispersión) de la nube (N) se utiliza la inercia

$\bar{\mathbb{X}}$: centro de masa

$$I_g(N) = E((\mathbb{X} - g)^t(\mathbb{X} - g))$$
momento de inercia relativo al punto g

9



Teorema de Huygens

$$\boxed{I_g(N) \geq I_{\bar{\mathbb{X}}}(N) = \Sigma_{\mathbb{X}}}$$

Poniendo el centro en $\bar{\mathbb{X}}$, la nube de puntos tiene la menor deformación

La inercia es mínima

10

$$\boxed{\Sigma_{\mathbb{X}}}$$

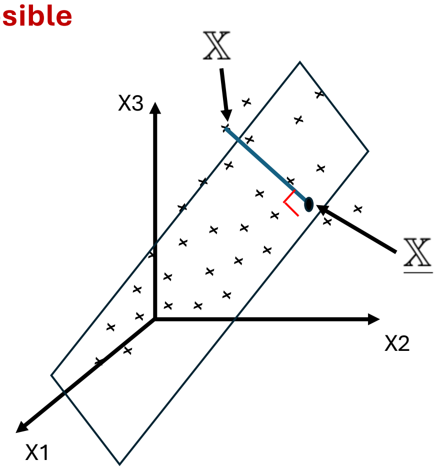
La matriz de varianzas y covarianzas mide la inercia y representa la información contenida en los datos (sus diferencias, semejanzas y relaciones)

11

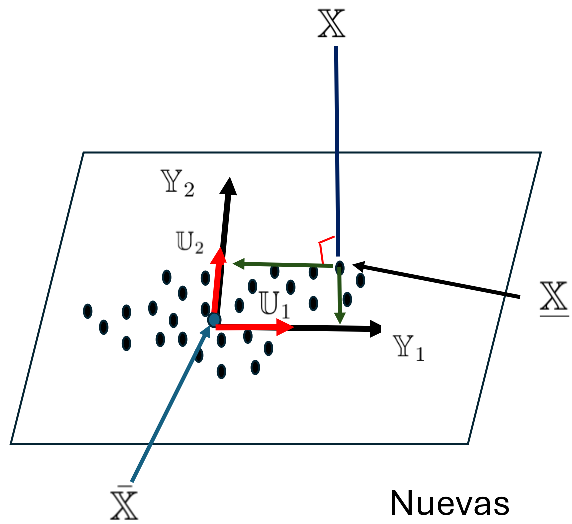
¿Cómo encontrar la base y reducir las dimensiones?

12

Idea:
Proyectar la nube de puntos en un subespacio vectorial centrado en una posición conveniente (espacio afín) con una dimensión menor.
Deformando la nube lo menos posible



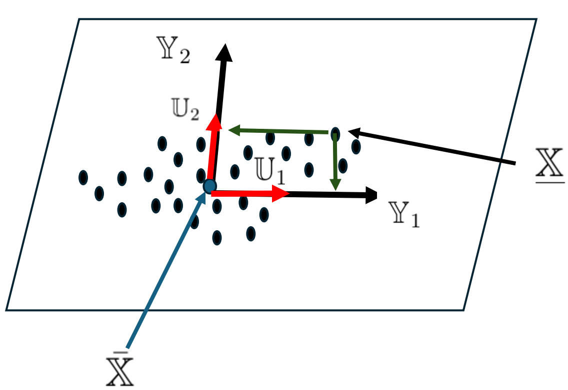
Proyecciones



Nuevas coordenadas (y_1, y_2)

13

Idea:
Proyectar la nube de puntos en un subespacio vectorial centrado en una posición conveniente (espacio afín) con una dimensión menor.
Deformando la nube lo menos posible



Observe que (de nuevo)

$$\boxed{\underline{\mathbb{X}} = g + \lambda_1 \mathbb{U}_1} + \lambda_2 \mathbb{U}_2$$

En general se busca una base $(\mathbb{U}_i)_{i=1,..n}$

tal que la nube proyectada se descomponga en:

$$\bar{\mathbb{X}} + \Re\{\mathbb{U}_1, ..., \mathbb{U}_q\}$$

14

Nuestro punto de partida

$$\mathbb{X} = (X_1, ..X_p)^t$$

$\Sigma_{\mathbb{X}}$ definida positiva

Valores propios: $\lambda_1 \geq \lambda_2 \geq .... \geq \lambda_p$

Vectores propios: $V_1 \;,\; V_2 \;,\; .... \;,\; V_p$

Recuerde que, para cada i:

$$\Sigma_{\mathbb{X}} V_i = \lambda_i V_i$$

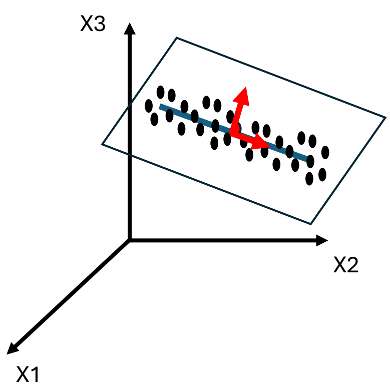
Cuando las escalas de medición son muy diferentes se trabaja con variables estandarizadas y entonces

$$\Sigma_{\mathbb{X}}$$

Es igual a la matriz de correlaciones

15

Idea:
Proyectar la nube de puntos en un subespacio vectorial centrado en una posición conveniente (espacio afín) con una dimensión menor.
Deformando la nube lo menos posible

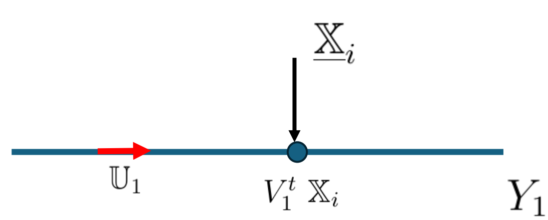


Cada componente (las coordenadas que determinan en cada eje) deben tener la mayor variabilidad posible (eso representa información): Varianza máxima

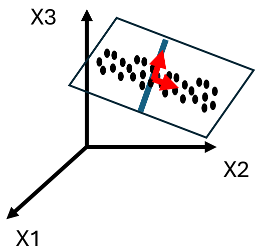
Primera componente:

$$\boxed{\begin{array}{l} \text{Tomando: } \| V_1 \| = 1 \\ Y_1 \;=\; V_1^t \;\mathbb{X} \\ var(Y_1) \;=\; \lambda_1 \end{array}}$$

Primer eje principal



16



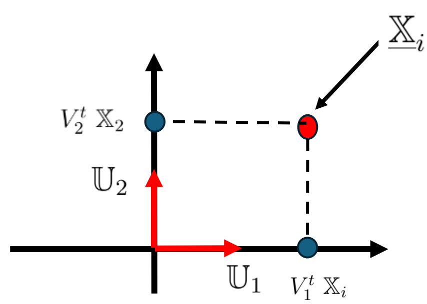
Primera componente:

$$\text{Tomando: } \| V_1 \| = 1$$
$$Y_1 = V_1^t \, \mathbb{X}$$
$$var(Y_1) = \lambda_1$$

Segunda componente:

$$\text{Tomando: } \| V_2 \| = 1$$
$$Y_2 = V_2^t \, \mathbb{X}$$
$$var(Y_2) = \lambda_2$$
$$cov(Y_1, Y_2) = 0$$

Representación del individuo i (su proyección) en el 1er. plano principal [Y1 , Y2]



17

En resumen se tiene:

(1) Las variables originales

$$\mathbb{X} = (X_1, ..X_p)^t$$

(2) Las nuevas variables

$$\mathbb{Y} = (Y_1, ..Y_q)^t \; (q \leq p)$$
$$\text{con } Y_i = V_i^t \, \mathbb{X}$$

Que representan las coordenadas en un nuevo espacio donde se pueden representar los individuos

Las componentes principales son vectores ortogonales

Primera componente:

$$\text{Tomando: } \| V_1 \| = 1$$
$$Y_1 = V_1^t \, \mathbb{X}$$
$$var(Y_1) = \lambda_1$$

Segunda componente:

$$\text{Tomando: } \| V_2 \| = 1$$
$$Y_2 = V_2^t \, \mathbb{X}$$
$$var(Y_2) = \lambda_2$$
$$cov(Y_1, Y_2) = 0$$

Componente i:

$$\text{Tomando: } \| V_i \| = 1$$
$$Y_i = V_i^t \, \mathbb{X}$$
$$var(Y_i) = \lambda_i$$
$$\text{Para j} < \text{i (los anteriores):}$$
$$cov(Y_i, Y_j) = 0$$

18

Obtención de las componentes y las variables

19

Se desea analizar los resultados en 6 exámenes (valor de 0 a 10) de un grupo de 10 estudiantes.

(a) Caracterizar su desempeño

(b) Valorar la importancia de los exámenes en el contexto de los resultados

(c) Identificar si se observan semejanzas o diferencias entre estudiantes.

(d)¿Se podrá hacer la caracterización utilizando un conjunto menor de variables? En ese caso ¿qué significarían?

Archivo: examenes

```
> Ex
       mat fis esp hist  ef
Lucia  7.0 6.5 9.2  8.6 8.0
Pedro  7.5 9.4 7.3  7.0 7.0
Ines   7.6 9.2 8.0  8.0 7.5
Luis   5.0 6.5 6.5  7.0 9.0
Andres 6.0 6.0 7.8  8.9 7.3
Ana    7.8 9.6 7.7  8.0 6.5
Carlos 6.3 6.4 8.2  9.0 7.2
Jose   7.9 9.7 7.5  8.0 6.0
Sonia  6.0 6.0 6.5  5.5 8.7
Maria  6.8 7.2 8.7  9.0 7.0
```

20

Paso 1: Obtener las CP

Variables estandarizadas porque es lo más general

```
> #paquete FactoMineR
> library(FactoMineR)
>
> #obtención de las CP y las deposito
> #en un objeto
>
> CP_Ex=PCA(Ex,scale.unit=TRUE,ncp=5,graph=FALSE)
>
> #cuando se estandarizan las variables scale.unit=TRUE
> #ncp: número de componentes a analizar
```

Paso 2: Para los análisis utilizar el paquete factoextra

```
#para analizar se utiliza el
#paquete factoextra

library(factoextra)
```

21

```
> #Las variables
> #la funcion get_pca_var brinda varias info
>
> var_Ex=get_pca_var(CP_Ex)
>
> #se depositan en este objeto, sus diferentes componentes
> #contienen diversos criterios
> #caso donde se obtienen las  coordenadas
> #es decir, los vectores propios que constituyen
> #los coeficientes de las CP
>
> #visualizacion de las coordenadas
> var_Ex$coord
```

Obtención de las componentes principales (dimensiones, variables sintéticas)

Vectores propios

```
> var_Ex$coord
          Dim.1      Dim.2       Dim.3       Dim.4       Dim.5
mat   0.8957980 -0.3452036  0.25797931 -0.09146818  0.05882803
fis   0.7227976 -0.6483946  0.02384033  0.23587773 -0.03068234
esp   0.6108931  0.7173206  0.33102532 -0.02454152 -0.04561456
hist  0.5999227  0.7484701 -0.23206345  0.15639747  0.03964443
ef   -0.9139265  0.1196373  0.34065108  0.18315368  0.02892890
```

22

Componentes (Vectores propios)

```
> var_Ex$coord
          Dim.1      Dim.2       Dim.3       Dim.4       Dim.5
mat   0.8957980 -0.3452036  0.25797931 -0.09146818  0.05882803
fis   0.7227976 -0.6483946  0.02384033  0.23587773 -0.03068234
esp   0.6108931  0.7173206  0.33102532 -0.02454152 -0.04561456
hist  0.5999227  0.7484701 -0.23206345  0.15639747  0.03964443
ef   -0.9139265  0.1196373  0.34065108  0.18315368  0.02892890
```

Obtención de las varianzas respectivas **(Valores propios)**

Varianzas (Valores propios)

```
> #valores propios
> vp=get_eigenvalue(CP_Ex)
> vp
       eigenvalue variance.percent cumulative.variance.percent
Dim.1 2.893249673       57.8649935                    57.86499
Dim.2 1.628650425       32.5730085                    90.43800
Dim.3 0.346596049        6.9319210                    97.36992
Dim.4 0.122612460        2.4522492                    99.82217
Dim.5 0.008891393        0.1778279                   100.00000
```

23

Determinación de las componentes principales más importantes (con las que se va a trabajar)

24

(1) Las variables originales

$$\mathbb{X} = (X_1, ..X_p)^t$$

$$\Sigma_{\mathbb{X}}$$

(2) Las nuevas variables

$$\mathbb{Y} = (Y_1, ..Y_q)^t \ (q \leq p)$$

$$var(Y_i) \ = \ \lambda_i$$

Respecto a la variabilidad

$$tr(\Sigma_{\mathbb{X}}) \ = \sum_{i=1}^{p} \sigma_{ii} = \text{Var Total}$$

$$tr(\Sigma_{\mathbb{X}}) \ = \sum_{i=1}^{p} var(Y_i) = \sum_{i=1}^{p} \lambda_i$$

$$\text{Var Total} = \sum_{i=1}^{p} \lambda_i$$

En su conjunto las Componentes Principales representan la variabilidad total de las variables originales

25

De acá se puede caracterizar la importancia de los nuevos ejes (de las Componentes Principales)

$$\text{Para } i = 1, ..., n$$

$$var(Y_i) \ = \ \lambda_i$$

$$\text{Var Total} = \sum_{i=1}^{p} \lambda_i$$

Varianza (inercia) explicada por una componente (un eje)

$$\text{Para } i = 1, ..., n$$

$$\frac{\lambda_i}{Var\ Total}$$

Varianza (inercia) explicada por el plano $[Y_i , Y_j]$

$$\text{Para } i, j = 1, ..., n \ i \neq j$$

$$\frac{\lambda_i + \lambda_j}{Var\ Total}$$

Varianza (inercia) acumulada hasta la componente k ($\lambda_1 \geq \ldots \geq \lambda_q$)

$$\frac{\sum_{i=1}^{k} \lambda_i}{Var\ Total}$$

26

No hay direcciones privilegiadas

Si para todo i: $\frac{\lambda_i}{Var\ Total} \approx \frac{1}{p}$

Tomar q que asegure: inercia acumulada sea mayor a 80% o 90%

$$\boxed{\frac{\lambda_i}{Var\ Total}}$$

$$\frac{\sum_{i=1}^{k} \lambda_i}{Var\ Total}$$

Varianza (inercia) explicada por una componente (un eje)

Varianza (inercia) acumulada hasta la componente k

```
> vp
       eigenvalue variance.percent cumulative.variance.percent
Dim.1 2.893249673       57.8649935                    57.86499
Dim.2 1.628650425       32.5730085                    90.43800
Dim.3 0.346596049        6.9319210                    97.36992
Dim.4 0.122612460        2.4522492                    99.82217
Dim.5 0.008891393        0.1778279                   100.00000
```

Varianzas (Valores propios)

27

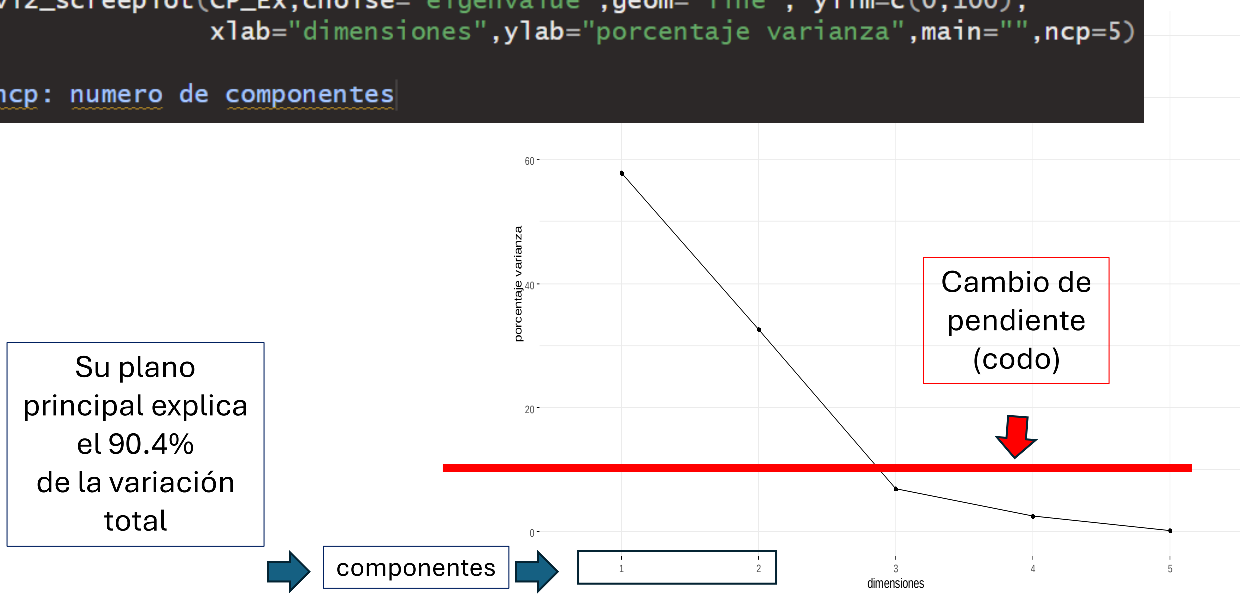
```
> vp
       eigenvalue variance.percent cumulative.variance.percent
Dim.1 2.893249673       57.8649935                    57.86499
Dim.2 1.628650425       32.5730085                    90.43800
Dim.3 0.346596049        6.9319210                    97.36992
Dim.4 0.122612460        2.4522492                    99.82217
Dim.5 0.008891393        0.1778279                   100.00000
```

Se toman dos ejes principales

Otra forma de valorar la cantidad de ejes es hacer un diagrama de pendiente (SCREE)

28

```
#diagrama de pendiente

fviz_screeplot(CP_Ex,choise="eigenvalue",geom="line", ylim=c(0,100),
              xlab="dimensiones",ylab="porcentaje varianza",main="",ncp=5)

#ncp: numero de componentes
```



29

En resumen

30

Datos originales

```
> Ex
       mat fis esp hist  ef
Lucia  7.0 6.5 9.2  8.6 8.0
Pedro  7.5 9.4 7.3  7.0 7.0
Ines   7.6 9.2 8.0  8.0 7.5
Luis   5.0 6.5 6.5  7.0 9.0
Andres 6.0 6.0 7.8  8.9 7.3
Ana    7.8 9.6 7.7  8.0 6.5
Carlos 6.3 6.4 8.2  9.0 7.2
Jose   7.9 9.7 7.5  8.0 6.0
Sonia  6.0 6.0 6.5  5.5 8.7
Maria  6.8 7.2 8.7  9.0 7.0
```

```
> var_Ex$coord
          Dim.1      Dim.2
mat   0.8957980 -0.3452036
fis   0.7227976 -0.6483946
esp   0.6108931  0.7173206
hist  0.5999227  0.7484701
ef   -0.9139265  0.1196373
```

```
> vp
       eigenvalue
Dim.1 2.893249673
Dim.2 1.628650425
```

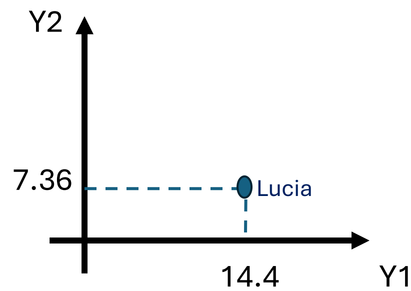
Coordenada de Lucia en el eje 1: 14.4

Componente 1 (eje principal 1):
Y1 = 0.89*mat+0.72*fis+0.61*esp+0.59*hist -0.91*ef

Coordenada de Lucia en el eje 2: 7.36

Componente 2 (eje principal 2):
Y2 = -0.34*mat-0.64*fis+0.71*esp+0.74*hist -0.11*ef

Representación de Lucia en el plano principal [Y1,Y2] (proyección)

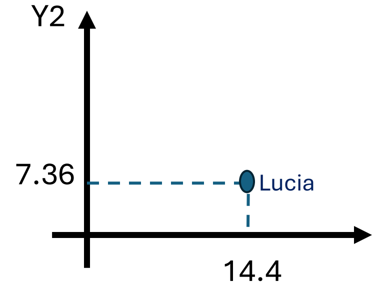


31

Interpretación de las componentes (ejes)
Aporte de las variables

32

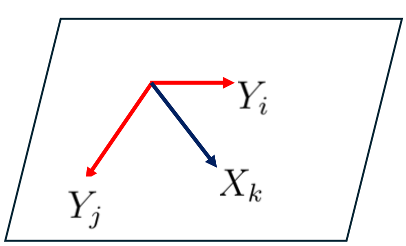
En las componentes principales los individuos se representan mediante sus coordenadas correspondientes



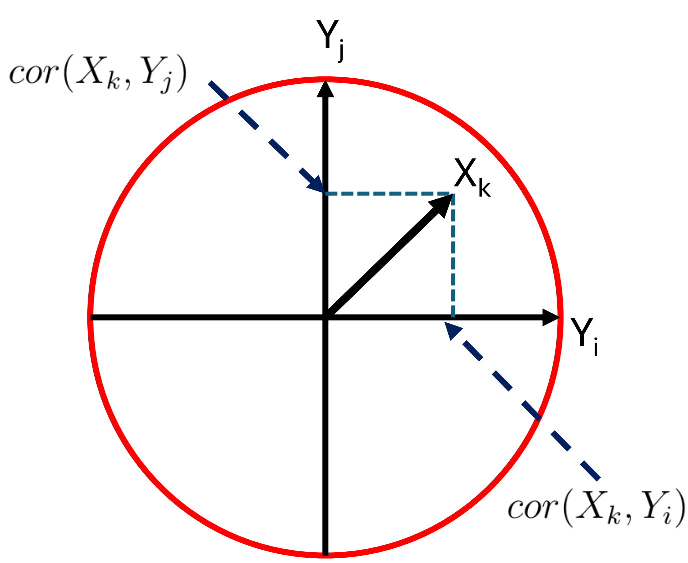
Para representar las variables y ver cómo influyen (se relacionan) sobre los ejes principales se utilizan las correlaciones

33

Plano principal [Y_i , Y_j]
Influencia de la variable X_k



Circulo de correlación



34

Volviendo al problema

Componente 1 (eje principal 1):
Y1 = 0.89*mat+0.72*fis+0.61*esp+0.59*hist -0.91*ef

Componente 2 (eje principal 2):
Y2 = -0.34*mat-0.64*fis+0.71*esp+0.74*hist -0.11*ef

$cor(mat, Y_1)$ $cor(mat, Y_2)$

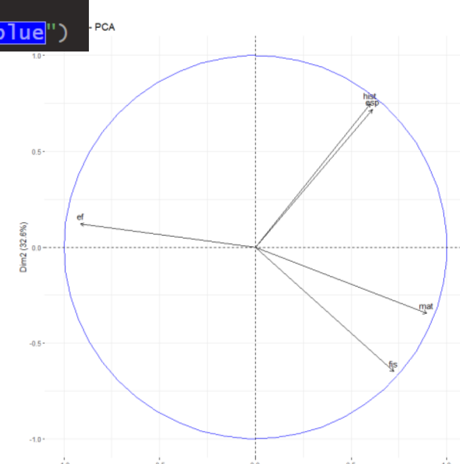
```
> #diagrama circulo de correlacion
> #se retoman los objetos que se encuentran en
> #var_Ex=get_pca_var(CP_Ex)
>
> #las correlaciones se encuentran en:
> var_Ex$cor
          Dim.1      Dim.2       Dim.3       Dim.4       Dim.5
mat   0.8957980 -0.3452036  0.25797931 -0.09146818  0.05882803
fis   0.7227976 -0.6483946  0.02384033  0.23587773 -0.03068234
esp   0.6108931  0.7173206  0.33102532 -0.02454152 -0.04561456
hist  0.5999227  0.7484701 -0.23206345  0.15639747  0.03964443
ef   -0.9139265  0.1196373  0.34065108  0.18315368  0.02892890
```

Para obtener las correlaciones

35

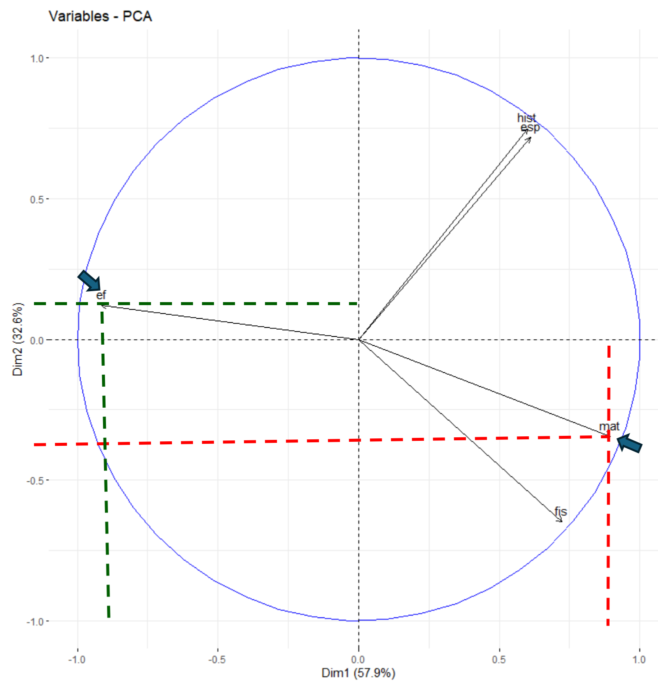
Circulo de correlación en el problema

```
#diagrama circulo de correlacion
#se retoman los objetos que se encuentran en
#var_Ex=get_pca_var(CP_Ex)
#las correlaciones se encuentran en:
var_Ex$cor

#el gráfico
fviz_pca_var(CP_Ex,axes=c(1,2),col.var="black",col.circle="blue")
```



36

```
> var_Ex$cor
          Dim.1      Dim.2
mat   0.8957980 -0.3452036
fis   0.7227976 -0.6483946
esp   0.6108931  0.7173206
hist  0.5999227  0.7484701
ef   -0.9139265  0.1196373
```

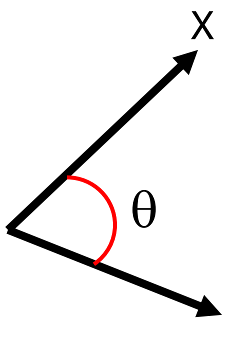


37

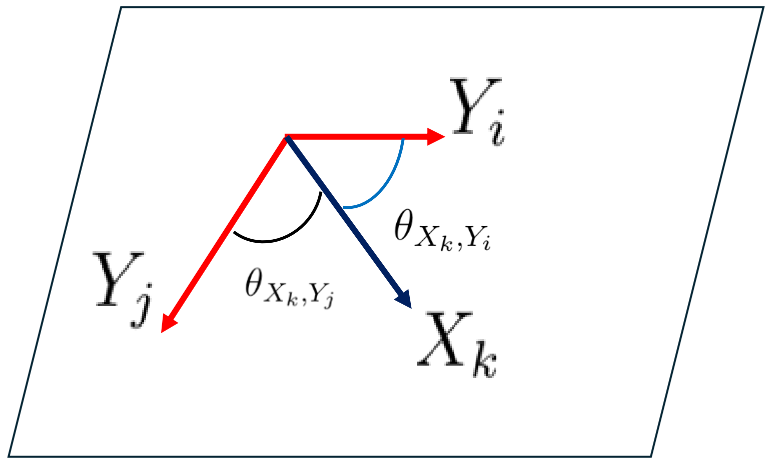
Otros componentes de análisis

38

Recordando:

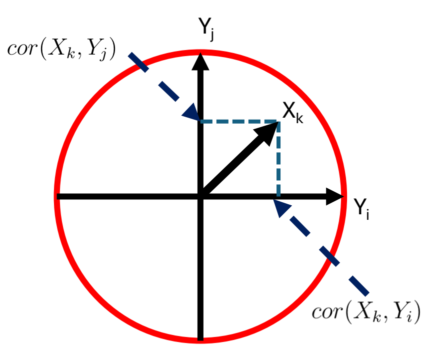


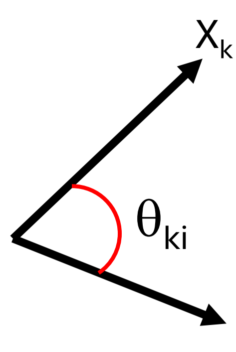
$$\mathrm{cor}(X,Y) = \cos(\theta)$$



39

Coseno del ángulo entre vectores





$$[\mathrm{cor}(X_k , Y_i)]^2 = \cos^2(\theta_{ki})$$

El cuadrado de los cosenos permite ver cercanía a los ejes $\cos^2(\theta_{ki}) \approx 1$ ($\theta_{ki} \approx 0$)

40

En el ejemplo

```
> #cuadrado de los cosenos
> #se retoman los objetos que se encuentran en
> #var_Ex=get_pca_var(CP_Ex)
> #los cuadrados de los cosenos se encuentran en:
>
> round(var_Ex$cos2,3)
```

$[cor(mat\ , Y_1)]^2 = cos^2(\theta_{11})$

```
> round(var_Ex$cos2,3)
     Dim.1 Dim.2 Dim.3 Dim.4 Dim.5
mat  0.802 0.119 0.067 0.008 0.003
fis  0.522 0.420 0.001 0.056 0.001
esp  0.373 0.515 0.110 0.001 0.002
hist 0.360 0.560 0.054 0.024 0.002
ef   0.835 0.014 0.116 0.034 0.001
```

41

Contribución de la variable al eje principal k (componente Y_k)

```
> round(var_Ex$cos2,3)
     Dim.1 Dim.2 Dim.3 Dim.4 Dim.5
mat  0.802 0.119 0.067 0.008 0.003
fis  0.522 0.420 0.001 0.056 0.001
esp  0.373 0.515 0.110 0.001 0.002
hist 0.360 0.560 0.054 0.024 0.002
ef   0.835 0.014 0.116 0.034 0.001
```

C[mat,Y1] =

```
> 100*0.802/(sum(var_Ex$cos2[,1]))
[1] 27.7197
```

$C[X_k, Y_i]$: contribución variable k al eje i

$$C[X_k, Y_i] = 100 * \frac{cos(\theta_{ik})}{\sum_{j=1}^{p} cos(\theta_{ij})}$$

42

```
> #contribucion de cada variable a los ejes principales
> #se retoman los objetos que se encuentran en
> #var_Ex=get_pca_var(CP_Ex)
```

```
> round(var_Ex$cos2,3)
     Dim.1 Dim.2 Dim.3 Dim.4 Dim.5
mat  0.802 0.119 0.067 0.008 0.003
fis  0.522 0.420 0.001 0.056 0.001
esp  0.373 0.515 0.110 0.001 0.002
hist 0.360 0.560 0.054 0.024 0.002
ef   0.835 0.014 0.116 0.034 0.001
```

C[mat,Y1] =

```
> 100*0.802/(sum(var_Ex$cos2[,1]))
[1] 27.7197
```

```
> round(var_Ex$contrib,1)
     Dim.1 Dim.2 Dim.3 Dim.4 Dim.5
mat   27.7   7.3  19.2   6.8  38.9
fis   18.1  25.8   0.2  45.4  10.6
esp   12.9  31.6  31.6   0.5  23.4
hist  12.4  34.4  15.5  19.9  17.7
ef    28.9   0.9  33.5  27.4   9.4
```

43

Los ejes y ef. Interpretación

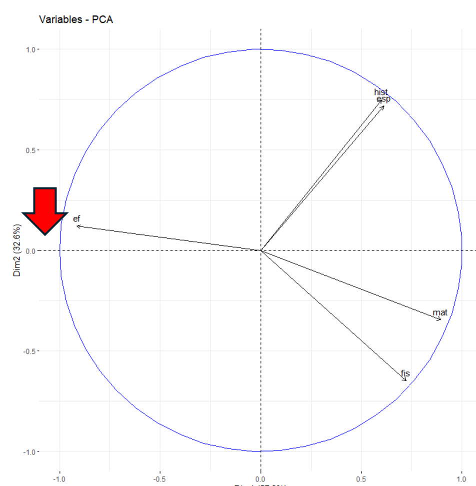
$[cor(ef\ ,\ Y_1)]^2 = \cos^2(\theta_{51})$

$[cor(ef\ ,\ Y_2)]^2 = \cos^2(\theta_{52})$

```
> round(var_Ex$cos2,3)
     Dim.1 Dim.2
mat  0.802 0.119
fis  0.522 0.420
esp  0.373 0.515
hist 0.360 0.560
ef   0.835 0.014
```

$C[X_k, Y_i]$: contribución variable k al eje i

```
     Dim.1 Dim.2
mat   27.7   7.3
fis   18.1  25.8
esp   12.9  31.6
hist  12.4  34.4
ef    28.9   0.9
```



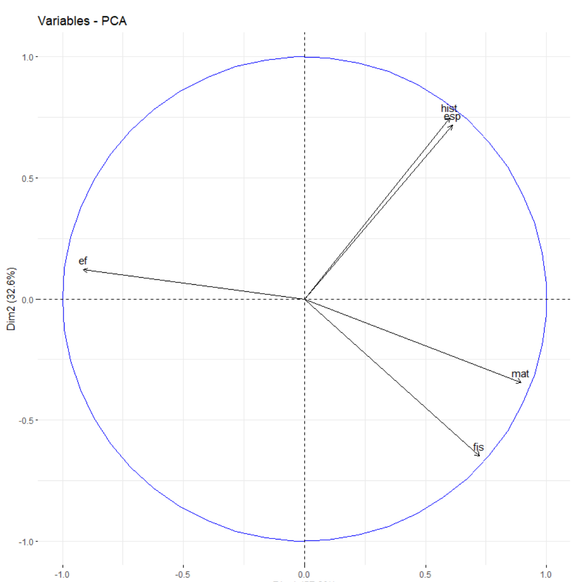

```
> var_Ex$cor
          Dim.1      Dim.2
mat   0.8957980 -0.3452036
fis   0.7227976 -0.6483946
esp   0.6108931  0.7173206
hist  0.5999227  0.7484701
ef   -0.9139265  0.1196373
```

44

Resumen: Caracterización de los ejes principales (1 y 2)

45

Componente 1 (eje principal 1):
Y1 = 0.89*mat+0.72*fis+0.61*esp+0.59*hist -0.91*ef
Varianza: 2.89 (57.8%)



Componente 2 (eje principal 2):
Y2 = -0.34*mat-0.64*fis+0.71*esp+0.74*hist -0.11*ef
Varianza: 1.6 (32.5%)

$[\mathrm{cor}(Y_i,X_k)]^2 = \cos^2(\theta_{ik})$

	Dim.1	Dim.2
mat	0.802	0.119
fis	0.522	0.420
esp	0.373	0.515
hist	0.360	0.560
ef	0.835	0.014

$C[X_k, Y_i]$: contribución variable k al eje i

	Dim.1	Dim.2
mat	27.7	7.3
fis	18.1	25.8
esp	12.9	31.6
hist	12.4	34.4
ef	28.9	0.9

46

Eje principal 1

Componente 1 (eje principal 1):
Y1 = 0.89*mat+0.72*fis+0.61*esp+0.59*hist -0.91*ef
Varianza: 2.89 (57.8%)

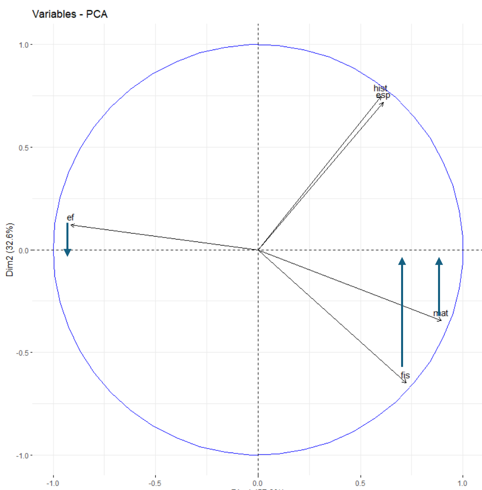
$[cor(Y_i,X_k)]^2 = cos^2(\theta_{ik})$

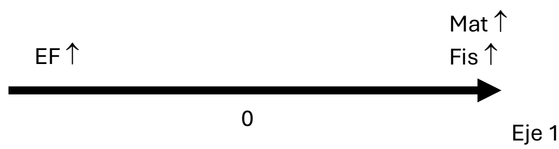
```
> cos(pi/4)^2
[1] 0.5
```

	Dim.1	Dim.2
mat	0.802	0.119
fis	0.522	0.420
esp	0.373	0.515
hist	0.360	0.560
ef	0.835	0.014

$C[X_k, Y_i]$: contribución variable k al eje i

	Dim.1	Dim.2
mat	27.7	7.3
fis	18.1	25.8
esp	12.9	31.6
hist	12.4	34.4
ef	28.9	0.9





47

Eje principal 2

Componente 2 (eje principal 2):
Y2 = -0.34*mat-0.64*fis+0.71*esp+0.74*hist -0.11*ef
Varianza: 1.6 (32.5%)

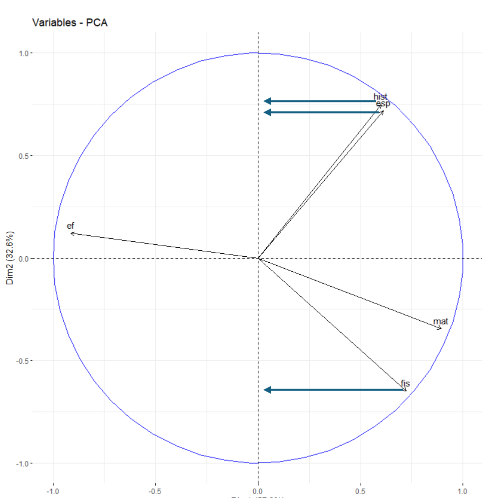
$[cor(Y_i,X_k)]^2 = cos^2(\theta_{ik})$

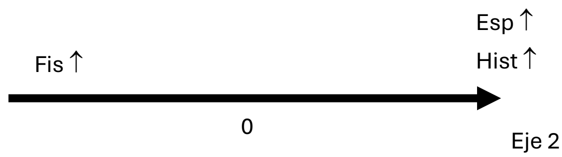
```
> cos(pi/4)^2
[1] 0.5
```

	Dim.1	Dim.2
mat	0.802	0.119
fis	0.522	0.420
esp	0.373	0.515
hist	0.360	0.560
ef	0.835	0.014

$C[X_k, Y_i]$: contribución variable k al eje i

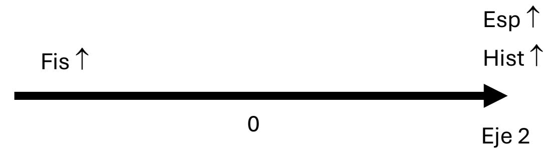
	Dim.1	Dim.2
mat	27.7	7.3
fis	18.1	25.8
esp	12.9	31.6
hist	12.4	34.4
ef	28.9	0.9



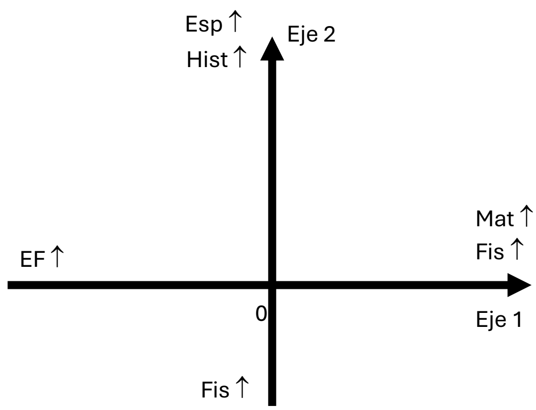


48

En resumen:



Para la representación de individuos



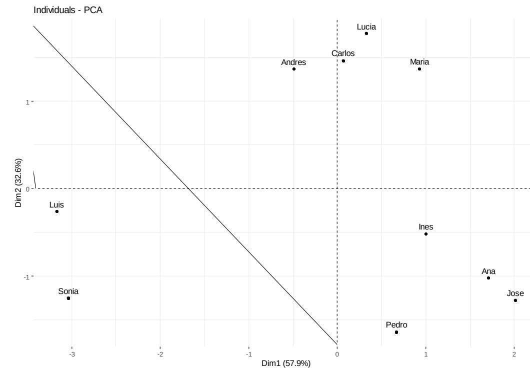
49

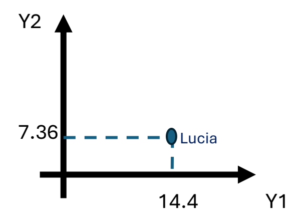
Comportamiento de los individuos

50

```
#Para representar los individuos

#se utiliza el comando:
#este comando tiene diversas informaciones
#sobre los individuos
ind=get_pca_ind(CP_Ex)

#las coordenadas se encuentran en:
ind$coord

#para graficar los individuos
fviz_pca_ind(CP_Ex,axes=c(1,2))
```

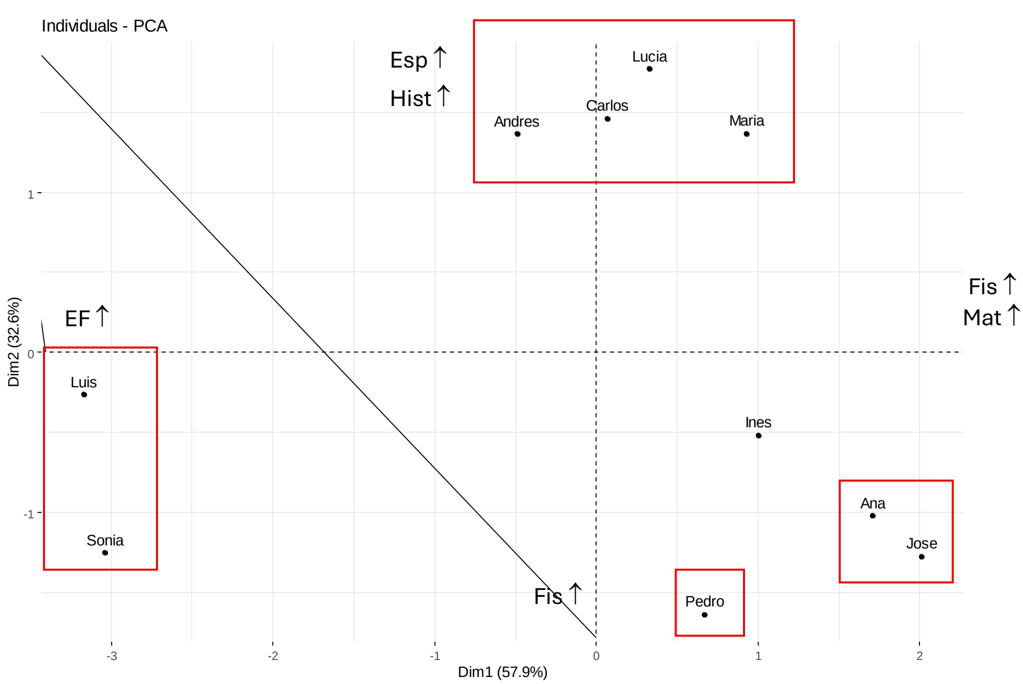
```
> ind$coord
             Dim.1      Dim.2
Lucia   0.32306263  1.7725245
Pedro   0.66544057 -1.6387021
Ines    1.00254705 -0.5156925
Luis   -3.17209481 -0.2627820
Andres -0.48886797  1.3654021
Ana     1.70863322 -1.0217004
Carlos  0.06758577  1.4623364
Jose    2.01185516 -1.2758646
Sonia  -3.04203029 -1.2548807
Maria   0.92386867  1.3693593
```





recordemos

51



52

Un complemento, es la calidad de la representación de los individuos en los ejes (planos) principales

53

Para cada individuo hay un ángulo entre sus valores en las variables y los ejes principales

```
> #contiene información del angulo (coseno cuadrado)
> ind$cos2
             Dim.1       Dim.2       Dim.3        Dim.4        Dim.5
Lucia  0.022270827 0.670420670 0.306659839 0.0006458478 2.816992e-06
Pedro  0.139905502 0.848430539 0.006686527 0.0001680781 4.809354e-03
Ines   0.514468899 0.136122895 0.202439714 0.1365196756 1.044882e-02
Luis   0.936851990 0.006429392 0.013583605 0.0427712757 3.637375e-04
Andres 0.084139511 0.656353715 0.245603703 0.0085448999 5.358172e-03
Ana    0.732686110 0.261979570 0.004052795 0.0011209894 1.605349e-04
Carlos 0.001892733 0.886081139 0.106192189 0.0057625700 7.136907e-05
Jose   0.673612108 0.270910359 0.048916504 0.0065104446 5.058468e-05
Sonia  0.808829929 0.137636943 0.017607237 0.0358004434 1.254472e-04
Maria  0.308554271 0.677869212 0.000310977 0.0018464085 1.141913e-02
```

$\cos^2(\theta_{ik}) \approx 1$ ($\theta_{ik} \approx 0$)

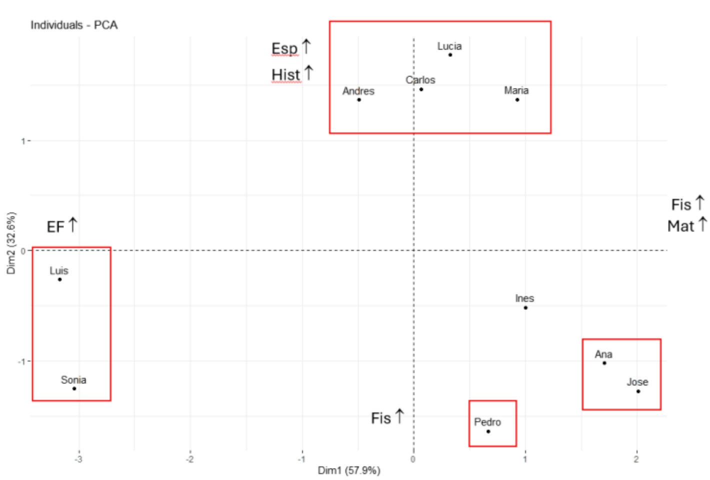
54

Para cada individuo también hay una contribución a cada uno de los ejes

```
> #tambien ind=get_pca_ind(CP_Ex)
> #contiene informacion sobre la contribucion
> #de cada individuo
> ind$contrib
             Dim.1      Dim.2       Dim.3       Dim.4       Dim.5
Lucia   0.36073437 19.2910834 41.46392357  0.24684974  0.01484748
Pedro   1.53049754 16.4881591  0.61060555  0.04338706 17.11987788
Ines    3.47395038  1.6328779 11.41096846 21.75259335 22.95871968
Luis   34.77814436  0.4239976  4.20932799 37.46613853  4.39379307
Andres  0.82603273 11.4470414 20.12771563  1.97950024 17.11709152
Ana    10.09047896  6.4094282  0.46591936  0.36428947  0.71941493
Carlos  0.01578791 13.1300601  7.39418080  1.13423412  0.19371414
Jose   13.98967133  9.9949649  8.48038057  3.19050613  0.34184774
Sonia  31.98461714  9.6688984  5.81215853 33.40593699  1.61421395
Maria   2.95008527 11.5134890  0.02481953  0.41656436 35.52647960
```

¡Inés está mal representada en el gráfico anterior!

55



Aplicación:
Ana y José se asemejan y se caracterizan por alto desempeño en mat y física, pero bajo en EF.
El eje 1 es el más importante para caracterizarlos, porque

	Dim.1	Dim.2
Lucia	0.36073437	19.2910834
Pedro	1.53049754	16.4881591
Ines	3.47395038	1.6328779
Luis	34.77814436	0.4239976
Andres	0.82603273	11.4470414
Ana	10.09047896	6.4094282
Carlos	0.01578791	13.1300601
Jose	13.98967133	9.9949649
Sonia	31.98461714	9.6688984
Maria	2.95008527	11.5134890

56