Estadística Aplicada III

Componentes Principales

Jorge de la Vega Góngora

Departamento de Estadística,
Instituto Tecnológico Autónomo de México

Semana 6: 19/21 de septiembre de 2018

Conceptos

Introducción

- Usualmente el Análisis de Componentes Principales (PCA) se utiliza como una técnica de visualización, reducción de dimensión y también como un paso intermedio en el análisis de datos como técnica exploratoria.
- Una vez obtenidas componentes principales (CP), se pueden aplicar otras técnicas de análisis multivariado a las CP como si fueran los datos originales.
- Para el PCA, **no** se requiere el supuesto de normalidad para los vectores aleatorios, pero si los datos son normales, entonces se puede hacer inferencia (hacer estimación, algunas pruebas de hipótesis e intervalos de confianza).

Planteamiento del problema I

- Dado un vector aleatorio $\mathbf{x} = (X_1, \ldots, X_p)$, se busca un conjunto de combinaciones lineales $Y_i = \mathbf{a}_i'\mathbf{x}$, $i = 1, 2, \ldots, p$ de tal forma que la variabilidad de k de ellos sea aproximadamente la variabilidad correspondiente al todo el vector, con $k << p$.
- Geométricamente, la transformación representa la selección de un nuevo sistema de coordenadas, obtenido por rotación y traslación del sistema original a ejes en donde se maximiza la varianza en cada dirección. Las primeras k componentes principales expanden un subespacio que contiene *la mejor visualización en k dimensiones*: una proyección o 'sombra' vista en la dirección de más información.
- Vimos en la presentación de normalidad que la *transformación de componentes principales* está dada por

$$\mathbf{y} = \mathbf{P}'(\mathbf{x} - \boldsymbol{\mu})$$

donde $\mathbf{P} = [\mathbf{e}_1 | \cdots | \mathbf{e}_p]$ es la matriz de eigenvectores de $\boldsymbol{\Sigma}$ tal que $\boldsymbol{\Sigma} = \mathbf{P}\Lambda\mathbf{P}'$, y que bajo dicha transformación, los contornos de la distribución de las variables transformadas se pueden expresar como

$$\mathbf{y}'\Lambda^{-1}\mathbf{y} = k^2.$$

- En caso de que así sea, habrá casi la misma información en las k combinaciones lineales como en las p variables.

Componentes Principales poblacionales I

- Sea $\mathbf{x}$ un vector aleatorio de $p \times 1$ con matriz de covarianzas $\mathbf{\Sigma}$ que tiene eigenvalores $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_p \geq 0$.
- Sean $\mathbf{a}_1, \ldots, \mathbf{a}_p \in \mathbb{R}^p$ los coeficientes de p combinaciones lineales $Y_i = \mathbf{a}_i\mathbf{x}$. De lo que hemos revisado sabemos que $\operatorname{cov}(Y_i, Y_j) = \mathbf{a}_i'\mathbf{\Sigma}\mathbf{a}_j$.
- Las CP son las combinaciones lineales no correlacionadas tales que $\text{Var}(Y_1), \ldots, \text{Var}(Y_p)$ es la máxima posible.
- Se agrega la restricción adicional de que $||\mathbf{a}_i|| = 1$, para no incrementar de forma arbitraria $\text{Var}(Y_i)$.

Problema de optimización

El problema consiste en resolver el siguiente sistema de problemas de optimización:

$$
\begin{aligned}
\max_{\mathbf{a}_1'\mathbf{a}_1=1} \mathbf{a}_1'\mathbf{\Sigma}\mathbf{a}_1 &= \text{Var}(Y_1) \\
\max_{\substack{\mathbf{a}_i'\mathbf{a}_i=1 \\ \mathbf{a}_i'\mathbf{\Sigma}\mathbf{a}_k=0, k<i}} \mathbf{a}_i'\mathbf{\Sigma}\mathbf{a}_i &= \text{Var}(Y_i) \quad i = 2, \ldots, n
\end{aligned}
$$

Componentes Principales poblacionales II

- Este problema ya lo hemos resuelto antes, cuando vimos optimización de formas cuadráticas. La solución es:

$$\mathbf{a}_i^* = \mathbf{e}_i \text{ y } \mathrm{Var}(Y_i) = \lambda_i$$

los vectores y valores propios de $\mathbf{\Sigma}$.

Solución alternativa (multiplicadores de Lagrange) I

- Una forma alternativa de resolver el problema de optimización con restricciones es utilizar los multiplicadores de Lagrange.
- Para $i = 1$, maximizamos la función

$$f(\mathbf{a}_1, \lambda) = \mathbf{a}_1' \mathbf{\Sigma} \mathbf{a}_1 - \lambda(\mathbf{a}_1'\mathbf{a}_1 - 1)$$

  donde λ es el multiplicador de Lagrange. Derivando respecto a $\mathbf{a}_1$ e igualando a 0 se obtiene el sistema:

$$\begin{aligned} \mathbf{\Sigma}\mathbf{a}_1 - \lambda\mathbf{a}_1 &= 0 \\ (\mathbf{\Sigma} - \mathbf{I}_p)\mathbf{a}_1 &= 0 \end{aligned}$$

  Por lo tanto, λ es un eigenvalor y $\mathbf{a}_1$ es un eigenvector. ¿Cuál de los eigenvectores? Notemos que:

$$\mathbf{a}_1'\mathbf{\Sigma}\mathbf{a}_1 = \mathbf{a}_1'\lambda\mathbf{a}_1 = \lambda||\mathbf{a}_1|| = \lambda$$

  Entonces la función se maximiza cuando $\lambda^* = \lambda_1$, el eigenvalor más grande, y en ee caso, $\mathbf{a}_1 = \mathbf{e}_1$.
- Para los subsecuentes casos, se agrega adicionalmente la restricción $\mathbf{a}_i\mathbf{a}_j = 0$ utilizando un segundo multiplicador y tomando las derivadas correspondientes:

$$f(\mathbf{a}_i, \lambda, \phi) = \mathbf{a}_i'\mathbf{\Sigma}\mathbf{a}_i - \lambda(\mathbf{a}_i'\mathbf{a}_i - 1) - \phi(\mathbf{a}_i'\mathbf{a}_j)$$

Observaciones I

- Noten que $\mathrm{tr}(\mathbf{\Sigma}) = \sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{n} \mathrm{Var}(Y_i)$, y entonces la proporción de la varianza total explicada o debida a la $j-$ésima componente principal es

$$\frac{\lambda_i}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} \quad i = 1, \ldots, p.$$

- Si la mayor parte (digamos, un 70 % a 90 %) de la varianza total se puede atribuir a las primeras cuantas componentes, entonces esas componentes pueden "sustituir" a las p variables sin mucha pérdida de información.
- Las componentes del eigenvector $\mathbf{e}_i = (e_{i1}, \ldots, e_{ip})$ ponderan la importancia de las variables originales en la $i-$ésima CP, es decir, son proporcionales a la correlación entre la $i-$ésima componente y la $k-$ésima variable:

Teorema

Si $Y_i = \mathbf{e}_i'\mathbf{x}$ son las CP, $i = 1, 2, \ldots, p$, entonces $\rho_{Y_i, X_k} = e_{ik} \frac{\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}$.

Demostración.

Observaciones II

Si $\mathbf{a}_k = \mathbf{1}_k$, entonces $X_k = \mathbf{a}_k'\mathbf{x}$. Entonces podemos escribir:

$$\text{cov}\left(X_k, Y_i\right) = \text{cov}\left(\mathbf{a}_k'\mathbf{x}, \mathbf{e}_i'\mathbf{x}\right) = \mathbf{a}_k'\boldsymbol{\Sigma}\mathbf{e}_i = \mathbf{a}_k'\lambda_i\mathbf{e}_i = \lambda_i\mathbf{a}_k'\mathbf{e}_i = \lambda_i e_{ik}$$

Entonces

$$\rho_{Y_i, X_k} = \frac{\text{cov}\left(Y_i, X_k\right)}{\sqrt{\text{Var}(Y_i)}\sqrt{\text{Var}(X_k)}} = \frac{\lambda_i e_{ik}}{\sqrt{\lambda_i}\sqrt{\sigma_{kk}}} = e_{ik}\frac{\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}$$

Por lo tanto, cada e_{ik} es proporcional a la correlación entre Y_i y X_k.

$\square$

Transformaciones de los datos para CP I

- Usualmente la matriz $\mathbf{X}$ se transforma a una matriz de covarianzas $\mathbf{\Sigma}$ o de correlaciones $\boldsymbol{\rho}$ que es lo único necesario para estimar las CP.
- Las CP pueden ser sensibles a outliers, por lo que se recomienda usar estimaciones robustas de $\mathbf{\Sigma}$.
- Sin embargo, el análisis de CP **no es invariante** ante cambios de escala, por lo que las estimaciones obtenidas de $\boldsymbol{\rho}$ o de $\mathbf{\Sigma}$ pueden ser diferentes, y no hay transformación directa fácil entre ellas.

Ejemplo (Mardia 1979)

Supongamos que se tienen tres variables dadas por:

- $X_1 =$ peso, dado en lbs
- $X_2 =$ altura, dado en pies
- $X_3 =$ edad (en años)

Supongamos que queremos realizar un cambio de escala de las variables a kg, cm y décadas, respectivamente.
Hay dos maneras de poder hacer el cambio de escala:

1. Multiplicar las variables por los factores de conversión (0.453592, 30.48 y 0.1) y hacer PCA en la matriz obtenida de las variables rescaladas.
2. Llevar a cabo el PCA en la matriz de covarianzas de las variables originales y multiplicar los elementos de los componentes relevantes por los respectivos factores de conversión.

Transformaciones de los datos para CP II

Entonces no se llega al mismo resultado por los dos métodos. En general, no se llegará al mismo resultado si se usa la matriz $\mathbf{S}$ o la matriz $\mathbf{R}$. Algunas de las diferencias son:

- El porcentaje de varianza explicado por las CP de $\mathbf{R}$ diferirá del porcentaje basado en $\mathbf{S}$.
- Los coeficientes de las CP de $\mathbf{R}$ difieren también de las de $\mathbf{S}$
- Aún si se expresan las componentes de $\mathbf{R}$ en términos de las variables originales, no serán iguales a las de $\mathbf{S}$

- En general, se recomienda:
  - Usar $\mathbf{S}$ cuando las unidades entre las variables son similares
  - Usar $\mathbf{R}$ cuando las varianzas en las variables originales son muy diferentes. En este caso, la interpretación de las CP puede ser más sencilla.Las CP de $\mathbf{R}$ sí son invariantes ante cambios de escala, porque lo es $\mathbf{R}$.

$\square$

- Cuando las variables tienen diferentes unidades y escalas, se usa $\boldsymbol{\rho}$ para hacer las variables comparables.
- Usualmente se busca como primera aproximación para la interpretación de las CP, ver qué coeficientes de la respectiva componente son mayores (positivos o negativos), y también pensar en los cambios de signo de las variables (lo que define *contrastes* entre variables).

¿Cuántas CP?

Sólo en el caso normal se puede usar una guía formal para determinar cuántas CP son significativas. En general, se tienen las siguientes guías:

- Retener las CP que acumulen un umbral, por ejemplo, 80 %, de la variación total.
- Retener las CP cuyos eigenvalores sean mayores que el promedio $\bar{\lambda}$. Para una matriz de correlación, este promedio es 1.
- Usar una gráfica de codo (*scree plot*): $\{i, \lambda_i\}$ y buscar el "doblez" natural entre valores grandes y valores pequeños de λ_i.
- Probar la significancia de los componentes 'grandes'.

Inferencia I

- La relación entre los eigenvalores de la matriz poblacional $\mathbf{\Sigma}$ y los de la matriz muestral $\mathbf{S}$ se da en el siguiente teorema:

Teorema (Anderson, 1963)

Sea $\mathbf{\Sigma} > 0$ con eigenvalores distintos y sea $\mathbf{U} \sim \frac{1}{m} W_p(\mathbf{\Sigma}, m)$ con descomposiciones $\mathbf{\Gamma\Delta\Gamma}'$ y $\mathbf{PLP}'$ respectivamente. Entonces:

$$\sqrt{m}(\mathbf{l} - \boldsymbol{\lambda}) \to \mathcal{N}_p\left(0, 2\Delta^2\right)$$

donde $\mathbf{l} = diag(\mathbf{L})$ y $\boldsymbol{\lambda} = diag(\mathbf{\Delta})$

- Para probar la significancia de componentes grandes, se prueba H_0: últimos k eigenvalores son pequeños e iguales $(\lambda_{p-k+1} = \lambda_{p-k+2} = \cdots = \lambda_p)$ La estadística de prueba (LRT) es:

$$u = \left(n - \frac{2p+11}{6}\right)\left(k\log(\bar{\lambda}) - \sum_{i=1}^{p}\log(\lambda_i)\right) \sim \chi^2_v$$

donde $v = \frac{1}{2}(k-1)(k+2)$ y $\bar{\lambda} = \sum_{i=p-k+1}^{p} \frac{\lambda_i}{k}$

Inferencia II

- El procedimiento comienza evaluando $H_{02} : \lambda_{p-1} = \lambda_p$. Si se acepta, se evalúa $H_{03} : \lambda_{p-2} = \lambda_{p-1} = \lambda_p$, y así sucesivamente hasta que se rechace la primera hipótesis.

Herramientas en R I

- hay métodos de componentes principales en varios paquetes de R:
  - La función princomp calcula las componentes principales, usando una matriz de datos $\mathbf{X}$ o directamente las matrices de correlación o covarianza. El argumento cor controla si se usa correlación o covarianza, y también se le pueden pasar estimaciones robustos de la varianza.
  - FactoMiner: Exploratory data analysis methods to summarize, visualize and describe datasets.
  - ade4: Tools for multivariate data analysis.
  - ca: Computation and visualization of simple, multiple and joint correspondence analysis.
  - MASS: Functions and datasets to support Venables and Ripley, “Modern Applied Statistics with S"(4th edition, 2002).
  - Exposition: descriptive (i.e., fixed-effects) multivariate analysis with the singular value decomposition.
  - factoextra que permite en general visualizar y extraer información de análisis de datos exploratorio multivariado.

Aplicaciones

Ejemplo 1: Creación de un Score para evaluación (Mardia) I

- Se tienen las calificaciones de 88 estudiantes en diferentes materias (algunos exámenes son a libro abierto y otros a libro cerrado). Las variables son:
  - `MC`: Mecánica (libro cerrado)
  - `VC`: Álgebra lineal (libro cerrado)
  - `LO`: Álgebra moderna (libro abierto)
  - `NO`: Análisis (libro abierto)
  - `SO`: Estadística (libro abierto)

```
library(factoextra)

Loading required package: ggplot2
Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa

X <- read.table("../data/Mardia/openclosedbook.dat",header=T)
head(X)

  MC VC LO NO SO
1 77 82 67 67 81
2 63 78 80 70 81
3 75 73 71 66 81
4 55 72 63 70 68
5 63 63 65 70 63
6 53 61 72 64 73
```

- Nos interesa, a partir de las 5 evaluaciones, una evaluación global que preserve en la mayor medida posible la variabilidad de grupo de estudiantes y poderlos separar adecuadamente por su desempeño.

Ejemplo 1: Creación de un Score para evaluación (Mardia) II

- Una posibilidad es hacer un promedio de las observaciones, pero así se ponderan del mismo modo todas las variables que conforman la calificación sin importar su variabilidad.
- Aplicando la descomposición en componentes principales:

```
z <- princomp(X,cor=T)  #función de componentes principales
summary(z)

Importance of components:
                         Comp.1    Comp.2     Comp.3     Comp.4    Comp.5
Standard deviation     1.783530 0.8599836 0.66705706 0.62281006 0.4965788
Proportion of Variance 0.636196 0.1479144 0.08899303 0.07757848 0.0493181
Cumulative Proportion  0.636196 0.7841104 0.87310342 0.95068190 1.0000000
```

- De acuerdo a los resultados anteriores, la primera componente explica cerca del 62 % de la variabilidad total, mientras que las primeras dos componentes cubren el 80 %.
- para obtener las combinaciones lineales a cada componente, podemos ver los ponderadores (loadings):

Ejemplo 1: Creación de un Score para evaluación (Mardia) III

```
summary(z, loadings=T)

Importance of components:
                          Comp.1    Comp.2     Comp.3     Comp.4    Comp.5
Standard deviation     1.783530 0.8599836 0.66705706 0.62281006 0.4965788
Proportion of Variance 0.636196 0.1479144 0.08899303 0.07757848 0.0493181
Cumulative Proportion  0.636196 0.7841104 0.87310342 0.95068190 1.0000000

Loadings:
   Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
MC  0.400  0.645  0.621  0.146  0.131
VC  0.431  0.442 -0.705 -0.298  0.182
LO  0.503 -0.129         0.109 -0.847
NO  0.457 -0.388 -0.136  0.666  0.422
SO  0.438 -0.470  0.313 -0.659  0.234
```

Entonces las combinaciones lineales de las componentes principales son:

$$
\begin{aligned}
Z_1 &= +0.4MC + 0.431VC + 0.503LO + 0.457NO + 0.438SO \\
Z_2 &= +0.645MC + 0.442VC - 0.129LO - 0.388NO - 0.47SO \\
Z_3 &= +0.621MC - 0.705VC - 0.037LO - 0.136NO + 0.313SO \\
Z_4 &= +0.146MC - 0.298VC + 0.109LO + 0.666NO - 0.659SO \\
Z_5 &= +0.131MC + 0.182VC - 0.847LO + 0.422NO + 0.234SO
\end{aligned}
$$

Las componentes tienen varianzas dadas por: $3.18, 0.74, 0.44, 0.39, 0.25$, respectivamente.

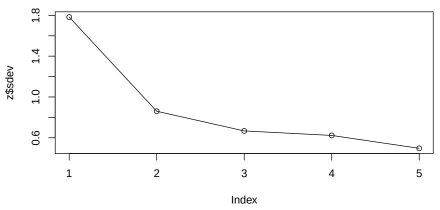
- Interpretación de las componentes:

Ejemplo 1: Creación de un Score para evaluación (Mardia) IV

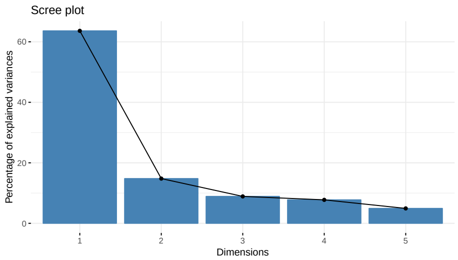
- la primera da un peso positivo a cada variable, por lo que representa una calificación promedio ponderada de acuerdo a la variabilidad de cada evaluación.
- La segunda componente se puede interpretar como un contraste entre las calificaciones a libro abierto y las que son a libro cerrado.

- La gráfica de codo para estos datos está dada a continuación. En esta gráfica no está claro dónde se da el doblez, que podría ser en la segunda o tercera componente. Vemos dos versiones de la gráfica de codo:

```r
plot(z$sdev,type="o")
```



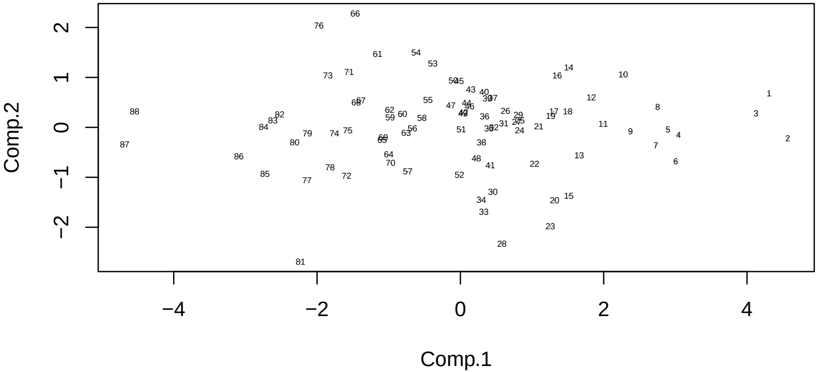

```r
fviz_screeplot(z)
```



- Podemos graficar las dos primeras componentes principales y asociar las evaluaciones a los diferentes alumnos para visualizar su distribución en la dirección de esas componentes.

Ejemplo 1: Creación de un Score para evaluación (Mardia) V

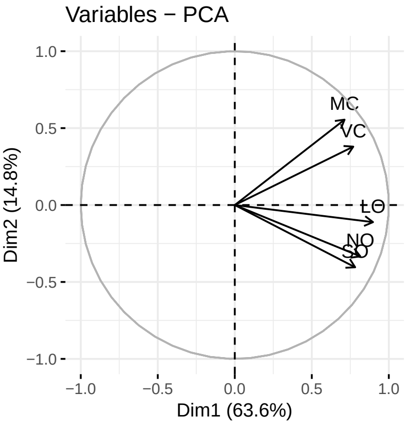
```r
plot(z$scores[,1:2],pch=16,cex=0.1)
text(z$scores,labels=1:88,cex=0.4) #Pon el número de alumno en el punto del i-ésimo score
```



Ejemplo 1: Creación de un Score para evaluación (Mardia) VI

- También podemos visualizar las CP como proyecciones en el espacio de las dos primeras componentes principales.

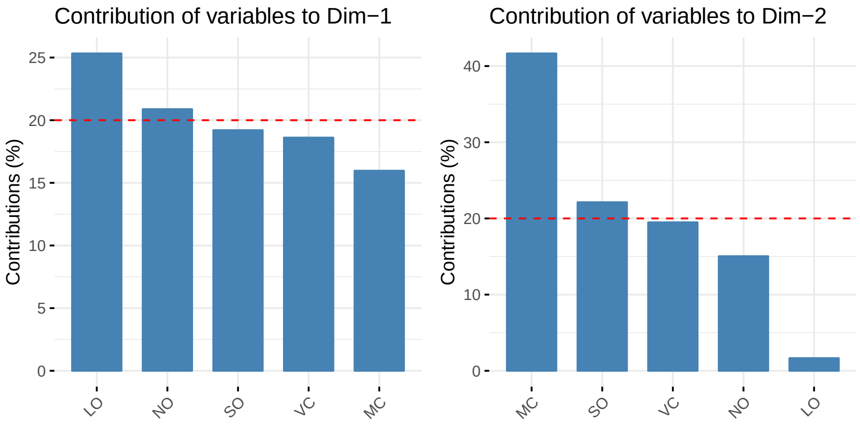
```
fviz_pca_var(z)
```



Ejemplo 1: Creación de un Score para evaluación (Mardia) VII

- También podemos ver la contribución de cada variable original a la CP:

```
library(gridExtra)
gr1 <- fviz_contrib(z,choice="var",axes=1)
gr2 <- fviz_contrib(z,choice="var",axes=2)
grid.arrange(gr1,gr2, ncol=2)
```



Ejemplo 1: Creación de un Score para evaluación (Mardia) VIII

- Suponiendo que los datos son aproximadamente normales, hagamos la serie de pruebas de hipótesis para los valores propios. Dado que $p = 5$ y $n = 88$,
  - Los valores de la estadística u para $i = 2, 3, 4, 5$ los calculamos con la fórmula dada, y los respectivos cuantiles:

```
p <- 5
u <- NULL
l <- z$sdev^2  #valores propios
for(i in 2:p)u[i-1] <- (88-(2*5+11)/6)*(i*log(mean(l[(p-i+1):p]))-sum(log(l)))
u

[1]   0.4329012 -64.6577141 -71.8823586 194.4604748

pchisq(u,df=0.5*(2:p-1)*(2:p+2),lower.tail=F)

[1] 8.053723e-01 1.000000e+00 1.000000e+00 7.417826e-34
```

Entonces, de acuerdo a las pruebas propuestas, nos quedamos con una componente (bajo el supuesto de que los datos son normales)

Ejemplo 2: Visualización de iris usando PCA I

- Podemos interpretar las componentes como proyecciones en las direcciones de máxima variabilidad de los datos.
- Las primeras CP son útiles para revelar estructura en los datos.

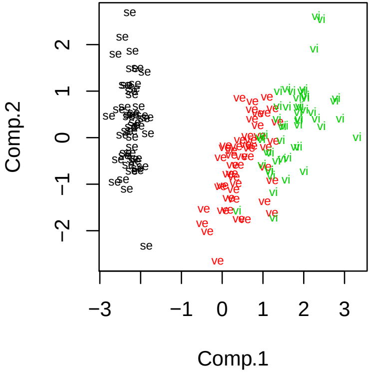
```
data("iris")
head(iris) #La última columna son las etiquetas de las variables

  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
4          4.6         3.1          1.5         0.2  setosa
5          5.0         3.6          1.4         0.2  setosa
6          5.4         3.9          1.7         0.4  setosa

iris.pc <- princomp(iris[,-5],cor=T)
summary(iris.pc,loadings=T)

Importance of components:
                          Comp.1    Comp.2     Comp.3      Comp.4
Standard deviation     1.7083611 0.9560494 0.38308860 0.143926497
Proportion of Variance 0.7296245 0.2285076 0.03668922 0.005178709
Cumulative Proportion  0.7296245 0.9581321 0.99482129 1.000000000

Loadings:
             Comp.1 Comp.2 Comp.3 Comp.4
Sepal.Length  0.521  0.377  0.720  0.261
Sepal.Width  -0.269  0.923 -0.244 -0.124
Petal.Length  0.580        -0.142 -0.801
Petal.Width   0.565        -0.634  0.524
```

Ejemplo 2: Visualización de `iris` usando PCA II

- Graficamos las primeras dos componentes. La primera componente separa claramente los datos en relación al ancho del pétalo. En este caso ya lo sabíamos por conocer previamente los datos, pero en general una CP puede ayudar a identificar cuando los datos forman grupos.

```
par(pty="s")
plot(iris.pc$scores[,1:2], type="n")
text(iris.pc$scores[,1:2],cex=0.6,labels=substr(iris[,5],1,2),
     col = as.numeric(factor(iris[,5])))
```

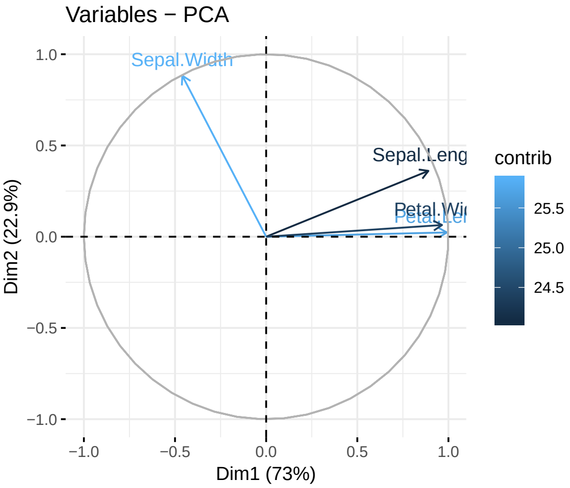
Ejemplo 2: Visualización de `iris` usando PCA III



Ejemplo 2: Visualización de `iris` usando PCA IV

- visualización de las componentes en el espacio de las dos primeras CP:

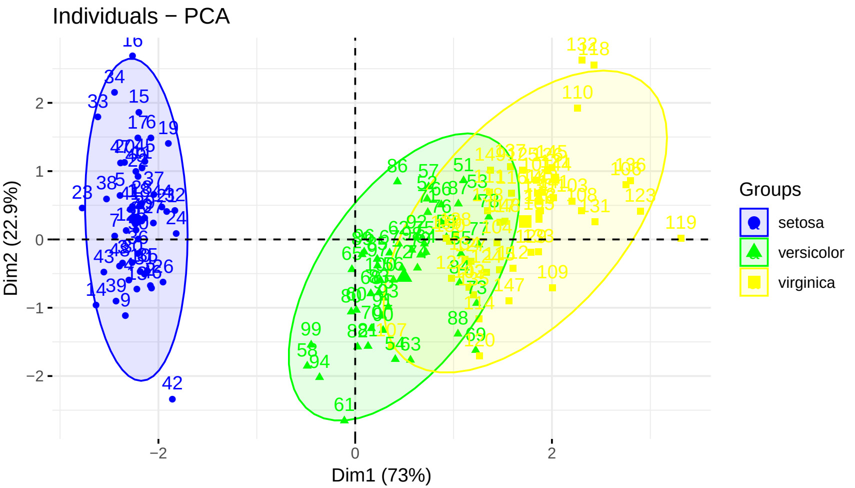
```
fviz_pca_var(iris.pc,col.var="contrib")
```



Ejemplo 2: Visualización de `iris` usando PCA V

- Podemos ver de otra manera las contribuciones de los diferentes grupos

```
fviz_pca_ind(iris.pc,habillage = iris[,5], palette=c("blue","green","yellow"),
         addEllipses = T)
```



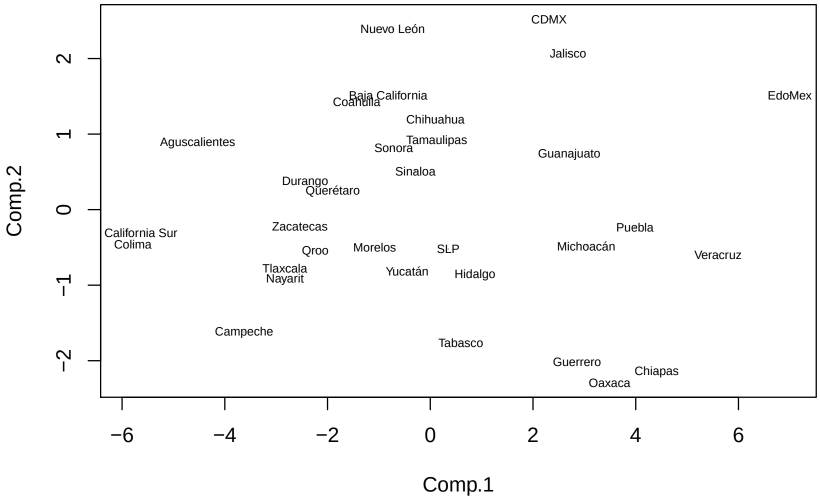
Ejemplo 3: Datos de pobreza del CONEVAL I

- A partir de los datos publicados por el CONEVAL en el Anexo de entodades federativas, se puede utilizar el PCA para reducir la dimensión de un problema.
- El archivo `ConevalPobreza2016.csv` resume la información para 32 entidades de la República Mexicana y se cuenta con 15 variables que corresponden a diferentes poblaciones en situaciones de pobreza.
- En este caso todas las variables están en las mismas unidades (miles de personas), por lo que se puede utilizar directamente la matriz de covarianzas.
- Se considerarán tres casos: los datos transformados a logaritmos, los datos originales usando covarianzas y los datos originales, usando correlaciones, para compara los resultados.

Ejemplo 3: Datos de pobreza del CONEVAL II

```
X <- read.csv("../data/ConevalPobreza2016.csv")
vars <- names(X)
vars #variables originales en el archivo

 [1] "Estado"
 [2] "Población.en.situación.de.pobreza"
 [3] "Población.en.situación.de.pobreza.moderada"
 [4] "Población.en.situación.de.pobreza.extrema"
 [5] "Población.vulnerable.por.carencias.sociales"
 [6] "Población.vulnerable.por.ingresos"
 [7] "Población.no.pobre.y.no.vulnerable"
 [8] "Población.con.al.menos.una.carencia.social"
 [9] "Población.con.al.menos.tres.carencias.sociales"
[10] "Rezago.educativo"
[11] "Carencia.por.acceso.a.los.servicios.de.salud"
[12] "Carencia.por.acceso.a.la.seguridad.social"
[13] "Carencia.por.calidad.y.espacios.en.la.vivienda"
[14] "Carencia.por.acceso.a.los.servicios.básicos.en.la.vivienda"
[15] "Carencia.por.acceso.a.la.alimentación"
[16] "Población.con.ingreso.inferior.a.la.línea.de.bienestar.mínimo"
[17] "Población.con.ingreso.inferior.a.la.línea.de.bienestar"

names(X) <- paste0("x",1:17) #cambio los nombres para hacerlos más manejables
x <- X[,-c(1:2)] #quitamos los nombres de los estados y una varriable que es combinación lineal de dos que ya es

pc.coneval <- princomp(log(x)) #se considera el log de los datos.
plot(pc.coneval$scores[,1:2],pch=16,cex=0.1)
text(pc.coneval$scores[,1:2],labels = X[,1],cex=0.6)
```

Ejemplo 3: Datos de pobreza del CONEVAL III



Ejemplo 3: Datos de pobreza del CONEVAL I

- Para interpretar las componentes, podemos ver los pesos de las variables:

```
options(width=120)
summary(pc.coneval,loadings = T)

Importance of components:
                          Comp.1    Comp.2     Comp.3      Comp.4      Comp.5      Comp.6      Comp.7      Comp.
Standard deviation     3.1153533 1.2747763 0.49164136 0.282307188 0.228391276 0.193604790 0.158762682 0.14646794
Proportion of Variance 0.8215432 0.1375573 0.02046033 0.006746207 0.004415448 0.003172839 0.002133598 0.00181593
Cumulative Proportion  0.8215432 0.9591005 0.97956083 0.986307036 0.990722484 0.993895323 0.996028921 0.99784485
                             Comp.9      Comp.10      Comp.11      Comp.12      Comp.13      Comp.14      Comp.1
Standard deviation     0.1032845533 0.0857215376 0.0695824723 0.0369712507 2.426869e-02 2.276924e-02 1.132292e-0
Proportion of Variance 0.0009029975 0.0006220076 0.0004098411 0.0001157029 4.985498e-05 4.388466e-05 1.085258e-0
Cumulative Proportion  0.9987478563 0.9993698639 0.9997797049 0.9998954078 9.999453e-01 9.999891e-01 1.000000e+0

Loadings:
    Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10 Comp.11 Comp.12 Comp.13 Comp.14 Comp.
x3   0.260         0.247         0.195  0.211  0.303  0.214         0.257   0.367   0.508   0.403           0.14
x4   0.361 -0.370  0.253 -0.108 -0.284 -0.430 -0.316  0.178 -0.131  0.435                           0.199
x5   0.175  0.276 -0.293         0.304 -0.109 -0.125  0.184  0.186  0.345  -0.619   0.299                  -0.10
x6   0.172  0.451  0.338  0.515 -0.225         0.227        -0.258  0.191  -0.208  -0.295          -0.217
x7   0.133  0.529 -0.225  0.178 -0.289 -0.282 -0.381  0.112  0.267 -0.192   0.407   0.128
x8   0.239  0.113        -0.114  0.151  0.101                0.233                 -0.249  -0.288           0.81
x9   0.293 -0.156 -0.246 -0.108                      -0.126         0.110   0.302          -0.106  -0.822
x10  0.254                              0.260 -0.263 -0.824         0.225                   0.102   0.205
x11  0.247  0.239 -0.144 -0.354         0.306 -0.248  0.181 -0.706 -0.202
x12  0.252               -0.214  0.185  0.198         0.245  0.370                 -0.619   0.200          -0.40
x13  0.267        -0.352 -0.184 -0.650  0.103  0.484         0.135        -0.209                   0.113
x14  0.313 -0.401 -0.365  0.676  0.165  0.201         0.149        -0.141                           0.161
x15  0.231        -0.119         0.375 -0.637  0.380 -0.240 -0.210 -0.238                   0.153   0.202
x16  0.313 -0.135  0.455                      -0.207         0.207 -0.591  -0.324   0.179   0.210  -0.239
x17  0.265         0.250                       0.175                        0.148   0.211  -0.780   0.159  -0.34
```

- Las dos primeras CP explican casi el 96 % de la variabilidad de los datos.

Ejemplo 3: Datos de pobreza del CONEVAL II

- La primera componente parece hacer un promedio ponderado sobre todas las variables de pobreza, generando un rankeo de las entidades de acuerdo a la multidimensionalidad de la pobreza.
- La segunda variable hace un contraste entre dos conjuntos de variables: $A = \{x_7, x_6, x_5\}$ que corresponde a la población que no es pobre y no es vulnerable, vulnerable por ingresos o vulnerable por carencias sociales, y el conjunto $B = \{x_{14}, x_4, x_9\}$ que son: carencia de servicios básicos de vivienda, pobreza extrema y al menos 3 carencias sociales. Entonces en conjunto parece que la segunda componente contrasta áreas urbanas con respecto a áreas rurales.
- La gráfica de codo nos da el tamaño de los eigenvalores:

```
plot(pc.coneval$sdev^2,type="o")
```

Ejemplo 3: Datos de pobreza del CONEVAL III