

Trabajo práctico N°2

Mariana Vargas V.

July 1, 2017

1 Análisis de Componentes Principales

El Análisis de Componentes Principales es una técnica multivariada que consiste en una transformación ortogonal de una matrix de datos en otra cuyas variables resulten linealmente independientes. Estas, llamadas *componentes principales*, son tales que representan las direcciones de mayor variabilidad de los datos, permitiendo así la posibilidad de reducir la dimensión.

1.1 Nuestros datos

Llevamos adelante el análisis de componentes principales en la base de datos `indice.RData` que tiene información sobre las finanzas de un grupo de empresas. Con el comando `prcomp` de R obtuvimos las componentes principales. La salida de `summary(pc)` sugiere que las dos primeras son las más importantes dado que representan el 99% de la varianza.

Importance of components:

	PC1	PC2	PC3	PC4
Standard deviation	2.7263	1.5827	0.19559	0.13514
Proportion of Variance	0.7433	0.2505	0.00383	0.00183
Cumulative Proportion	0.7433	0.9938	0.99759	0.99942

	PC5	PC6	PC7	PC8	PC9	PC10
	0.07218	0.01957	0.01399	0.006285	0.00238	2.239e-17
	0.00052	0.00004	0.00002	0.000000	0.00000	0.000e+00
	0.99994	0.99998	1.00000	1.000000	1.00000	1.000e+00

Podemos confirmar esto en la figura 1 en donde se muestran un gráfico de barra de las componentes contra la varianza.

En la figura 2 podemos ver a las variables representadas en función de las dos primeras componentes.

Observamos que algunas variables se aproximan a la primera componente, tales como la rentabilidad económica, el margen de explotación y el costo marginal de financiamiento, y otras que se aproximan a la segunda componente, como la inmovilización del activo. También hay variables que mantienen con las componentes una relación negativa, como la inmovilización del patrimonio, el pasivo, y la solvencia. La matriz de correlaciones (tabla 1) confirma esto.

Repetimos el análisis usando la matriz de correlación de los datos. La salida muestra que esta vez las cuatro primeras componentes suman el 95% de la correlación entre las variables, lo cual podemos ver en la figura 3.

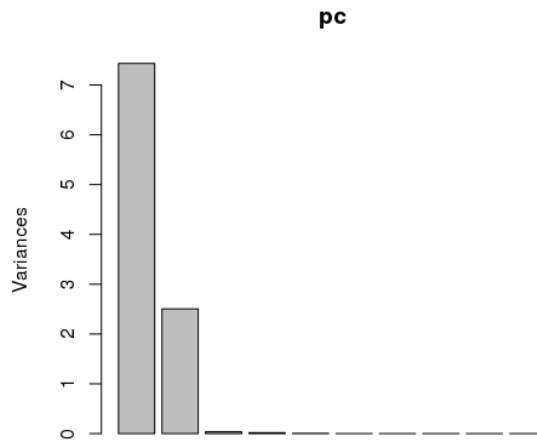


Figure 1: Componentes vs. la varianza.

Importance of components:

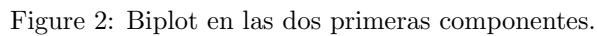
	PC1	PC2	PC3	PC4	PC5
Standard deviation	2.4670	1.5222	0.79441	0.71303	0.55569
Proportion of Variance	0.6086	0.2317	0.06311	0.05084	0.03088
Cumulative Proportion	0.6086	0.8403	0.90341	0.95425	0.98513
	PC6	PC7	PC8	PC9	PC10
	0.32151	0.16390	0.13310	0.02710	7.353e-18
	0.01034	0.00269	0.00177	0.00007	0.000e+00
	0.99547	0.99816	0.99993	1.00000	1.000e+00

La figura 4 muestra las variables originales en función de las componentes elegidas, que a diferencia del análisis anterior, maximizan la correlación entre variables y no su varianza.

De la matriz de correlación entre componentes y variables muestra una relación positiva entre la inmovilización del activo y del patrimonio, y del pasivo respecto de la primera componente, mientras que la propiedad del activo mantiene una relación negativa.

2 Análisis de correspondencia

Mediante el análisis de correspondencia estudiamos tablas de contingencia de variables clasificatorias. Esta técnica es análoga al análisis de componentes principales.



Trabajamos con una base de datos que tiene información sobre una encuesta sobre hogares. Analizamos las variables correspondientes al estado civil y al nivel de educación. Con el comando `ca` obtenemos el análisis de correspondencia. La raíz cuadrada de la suma de los autovalores nos da un índice de correlación entre filas y columnas. En este caso es de 0.4, lo que indica una baja correlación. Esto es, el nivel de educación no necesariamente define el estado civil, y viceversa. En la figura 5 podemos visualizar los resultados. Recordemos que puntos correspondientes a las filas que estén más cerca indican perfiles columna más parecidos. Por ejemplo, observar que las personas unidas y casadas tienden a tener el mismo nivel de educación, contrario a lo que ocurre con personas viudas y solteras, que se ven alejadas en el gráfico.

3

	1	2	3	4	5	6	7	8	9	10
LIQACID	-0.78	-0.62	0.03	0.01	-0.02	-0.01	-0.00	-0.00	-0.00	0.00
SOLVENC	-0.33	-0.94	0.10	0.03	-0.00	0.00	0.00	0.00	-0.00	0.00
PROPACT	0.94	-0.33	0.03	0.03	-0.00	0.01	0.00	-0.00	0.00	-0.00
PNOCOR	-0.97	0.21	-0.03	-0.01	-0.06	0.00	0.00	0.00	0.00	-0.00
AUTOFIN	1.00	0.08	0.04	-0.02	-0.01	-0.01	-0.00	0.00	0.00	-0.00
INMACT	-0.34	0.93	0.08	0.09	-0.00	-0.00	-0.00	0.00	-0.00	0.00
INMPN	-0.90	0.41	0.11	-0.08	0.01	0.00	-0.00	-0.00	0.00	-0.00
RENTECO	0.99	0.12	0.04	-0.01	-0.01	-0.00	0.00	-0.00	-0.00	-0.00
MAREXP	1.00	0.06	0.01	-0.02	-0.03	0.01	-0.01	0.00	0.00	0.00
REXP_INT	0.99	0.15	0.05	-0.01	-0.01	0.00	0.01	-0.00	-0.00	0.00

Table 1: Matriz de correlaciones entre componentes y variables usando la matriz S.

	1	2	3	4	5	6	7	8	9	10
LIQACID	-0.75	-0.65	0.04	0.01	0.00	0.12	0.05	0.04	0.02	0.00
SOLVENC	-0.81	-0.57	-0.06	0.04	-0.04	0.06	0.04	0.03	-0.02	0.00
PROPACT	-0.91	-0.28	-0.05	0.29	-0.01	-0.07	-0.04	-0.01	0.00	-0.00
PNOCOR	0.88	0.07	0.34	0.23	0.04	0.20	-0.05	0.03	-0.00	-0.00
AUTOFIN	-0.74	0.45	-0.18	-0.19	0.42	0.05	-0.03	0.04	-0.00	-0.00
INMACT	0.86	0.03	-0.11	0.44	0.21	-0.09	0.07	0.03	0.00	0.00
INMPN	0.94	0.08	-0.06	-0.30	-0.04	0.04	0.08	0.01	-0.00	-0.00
RENTECO	-0.72	0.65	0.05	0.17	0.02	0.11	0.07	-0.08	-0.00	-0.00
MAREXP	-0.67	0.40	0.61	-0.06	-0.00	-0.11	0.03	0.04	-0.00	-0.00
REXP_INT	-0.37	0.83	-0.29	0.12	-0.28	0.04	-0.00	0.07	0.00	0.00

Table 2: Matriz de correlación de las variables en función de las componentes.

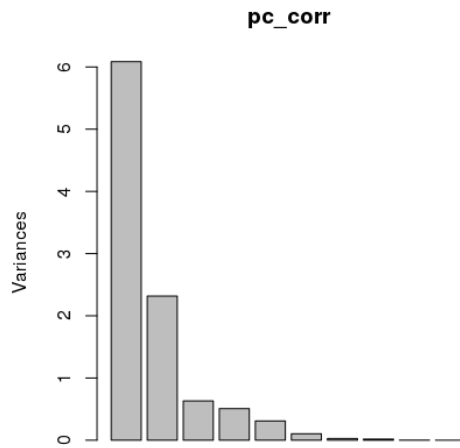


Figure 3: Gráfico de las componentes vs. la varianza, usando la matriz de correlación.

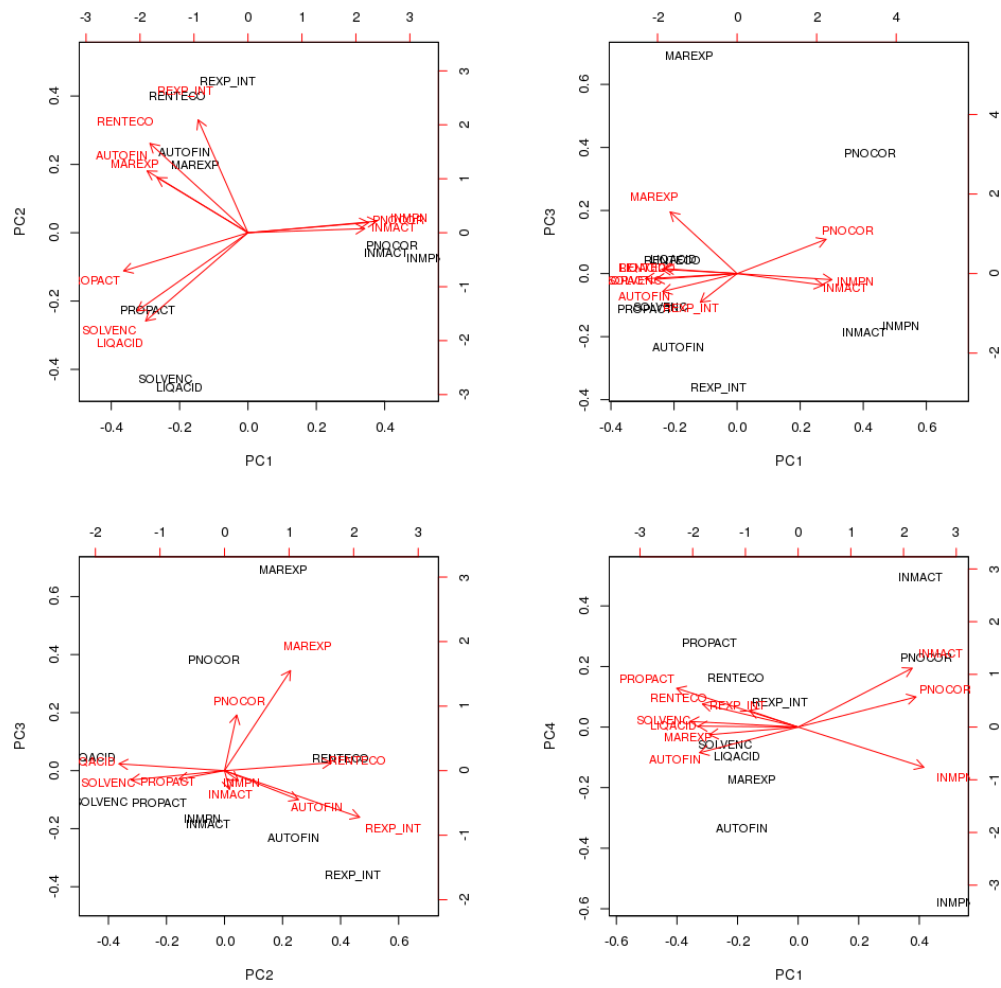


Figure 4: Gráfico de las variables en función de las componentes principales.

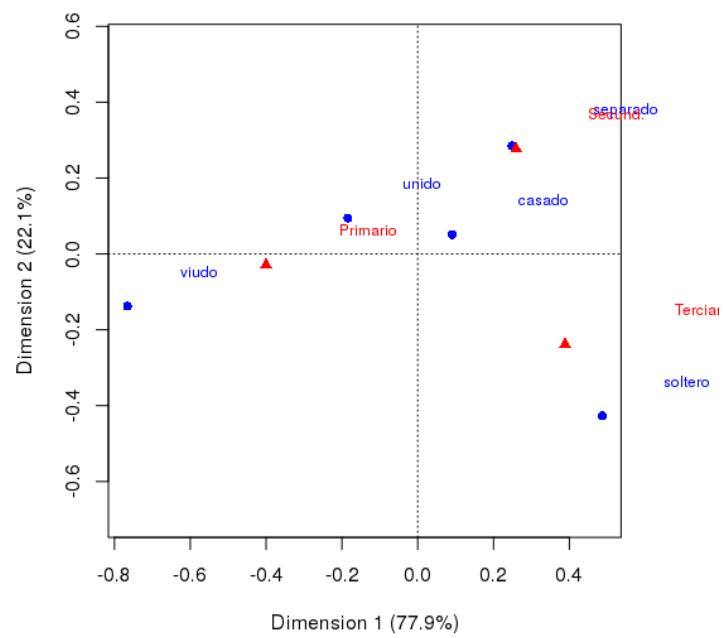


Figure 5: Gráfico de correspondencia simple entre estado civil y educación.

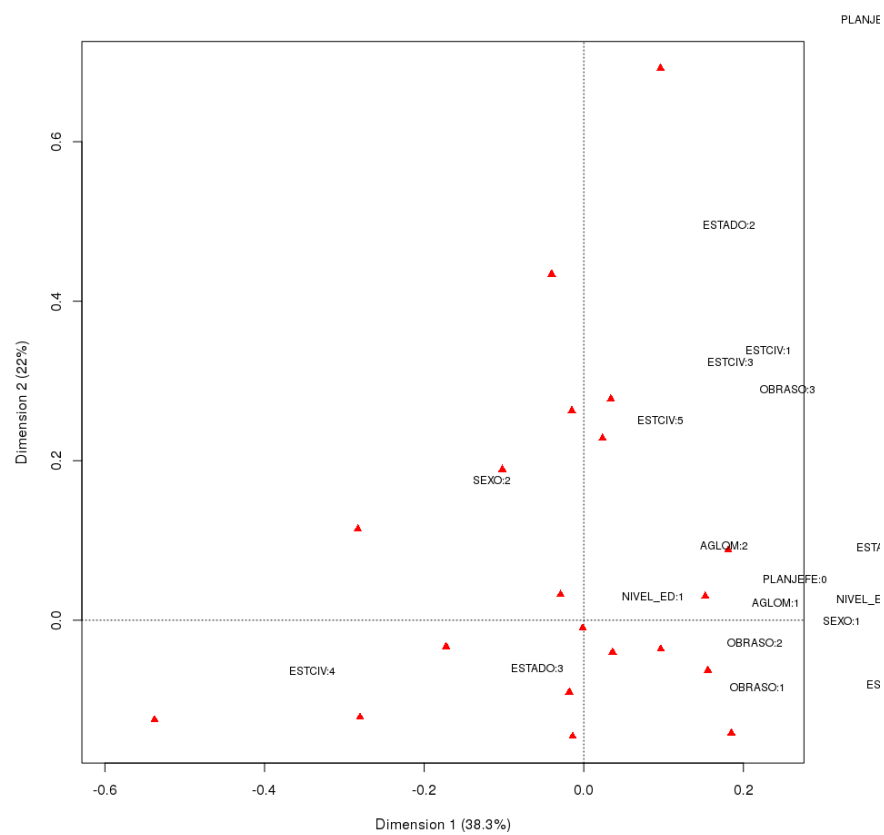


Figure 6: Gráfico de correspondencia simple entre estado civil y educación.