

Laboratorio 2: ACP con datos faltantes

KEVIN STEVEN GARCÍA^a, ALEJANDRO VARGAS^b

1. Introducción

En el presente informe veremos la aplicación del ACP con datos completos y datos faltantes a la matriz de datos correspondientes a las importaciones de los países sudamericanos provenientes de estados unidos entre 1991 y 2010. El ACP con datos faltantes se realizará a la matriz de datos mencionada con el 5,10,15, y 20 % de datos faltantes, por medio de dos algoritmos, el algoritmo NIPALS y el ACP-EM, se mostraran los resultados obtenidos y se comparará la eficiencia o la imputación obtenida de los dos algoritmos en cuanto a las estimaciones de los datos faltantes con respecto a la matriz de datos original. Además se analizará el comportamiento de los métodos en cuanto a valores y vectores propios obtenidos, porcentaje de Inercia explicado por los ejes, la estructura factorial en la nube de individuos y en la nube de variables, la ortogonalidad en las componentes y la ortonormalidad de los vectores propios obtenidos.

La base de datos sobre la cuál se va a trabajar es la siguiente:

Año	Colombia	Brasil	Chile	Argentina	Ecuador	Peru
1991	44.4	27.2	45.6	20.0	6.0	14.1
1992	75.5	11.8	58.9	22.6	17.8	14.4
1993	110.7	50.6	128.3	17.2	119.4	118.5
1994	80.3	70.6	102.2	15.2	154.9	146.1
1995	81.6	82.3	89.0	35.1	169.4	127.1
1996	76.4	97.4	185.0	51.0	75.5	129.0
1997	32.0	89.5	195.3	31.1	33.4	110.2
1998	55.5	63.1	66.3	24.4	9.7	66.7
1999	74.3	72.6	76.3	28.1	11.2	110.7
2000	84.5	76.2	80.1	29.5	11.8	110.2
2001	87.1	97.4	89.3	51.5	63.1	89.3
2002	89.3	89.5	72.4	40.3	66.3	70.2
2003	70.2	63.1	80.1	60.5	76.3	90.1
2004	90.1	66.3	70.5	39.1	20.0	64.5
2005	60.5	76.3	107.2	31.1	63.4	92.7
2006	140.3	20.0	63.4	50.2	101.2	120.8
2007	120.4	22.6	101.2	51.0	103.1	107.2
2008	130.2	17.2	103.1	42.5	66.7	70.8
2009	110.1	31.1	75.6	25.7	110.7	101.2
2010	120.2	24.4	68.9	60.3	110.2	110.8

^aCódigo: 1533173. E-mail: kevin.chica@correounivalle.edu.co

^bCódigo: 1525953. E-mail: jose.alejandro.vargas@correounivalle.edu.co

2. ACP datos completos

En el laboratorio 1 se vio paso por paso la aplicación del ACP, por lo tanto se mostraran solo los resultados obtenidos, es decir, los valores y vectores propios u , y las componentes principales Z .

Los valores y vectores propios obtenidos para la matriz de correlaciones asociada a la matriz de datos completa son:

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
2.1934092	1.9561781	0.9038789	0.5119470	0.2854407	0.1491461

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
-0.3266572	0.5708984	0.0068014	0.1170875	0.5332445	0.5189071
-0.1588587	-0.6033277	0.2009878	-0.5413708	0.1852719	0.4929051
-0.3381710	-0.4635960	0.0514265	0.7890499	-0.0779079	0.1985066
-0.2898593	0.2066273	0.8863988	-0.0448699	-0.2455986	-0.1589085
-0.5459391	0.1756044	-0.3785943	-0.2216216	-0.6626128	0.1990175
-0.6096157	-0.1470297	-0.1669623	-0.1395665	0.4193810	-0.6192861

Entonces, las componentes serán:

$$C = \begin{pmatrix} 3.57940293 & 0.2200186 & -0.45449358 & 0.07133223 & -0.720973151 & -0.20047529 \\ 2.98416620 & 1.1042398 & -0.46390509 & 0.71180698 & -0.455553845 & 0.19923050 \\ -1.18011107 & -0.0116174 & -1.74610834 & 0.70549306 & 0.310200546 & 0.48646527 \\ -1.55340630 & -0.7593625 & -2.19242499 & -0.62366337 & -0.204020462 & -0.20732998 \\ -1.75941910 & -0.3826535 & -0.86317511 & -1.17272788 & -0.865691407 & 0.12542483 \\ -1.95192515 & -2.1096857 & 1.13135115 & 0.91256447 & -0.033529896 & 0.19059358 \\ -0.24144158 & -3.3571269 & 0.21738043 & 1.42840546 & -0.256638938 & -0.32533218 \\ 1.98337243 & -0.7162408 & -0.17289755 & -0.37637158 & 0.202418839 & -0.25420678 \\ 0.73442430 & -0.7828704 & -0.07627906 & -0.46668505 & 1.061208739 & -0.52216426 \\ 0.53228046 & -0.6705084 & 0.04502957 & -0.41734440 & 1.233682980 & -0.25313938 \\ -0.37260533 & -0.5712985 & 1.32752250 & -0.83881402 & 0.043950908 & 0.54991653 \\ 0.34063969 & -0.2236844 & 0.59420850 & -0.93477773 & -0.009209529 & 0.85547004 \\ -0.25306113 & 0.1018299 & 1.54180034 & -0.54482210 & -0.818951578 & -0.47394187 \\ 1.13138487 & 0.1457486 & 0.74310059 & -0.28906437 & 0.448604684 & 0.38436376 \\ 0.26399863 & -1.2144046 & -0.14020651 & -0.12147196 & -0.240048044 & -0.04377519 \\ -1.30421117 & 2.4709380 & 0.21944271 & 0.01303094 & 0.504568106 & -0.33221632 \\ -1.21980212 & 1.6156676 & 0.38826932 & 0.71976837 & -0.148517945 & -0.21359085 \\ -0.07690509 & 1.8051120 & 0.27156821 & 1.24900452 & 0.206211966 & 0.49705097 \\ -0.35915931 & 1.2125308 & -1.25387195 & 0.04770938 & 0.035716022 & 0.03874664 \\ -1.27762217 & 2.1233680 & 0.88368884 & -0.07337295 & -0.293427995 & -0.50109002 \end{pmatrix}$$

3. ACP datos faltantes

3.1. NIPALS

El algoritmo NIPALS realiza una descomposición singular de una matriz de datos, mediante secuencias iterativas convergentes de proyecciones ortogonales, el cuál utiliza la descomposición derivada del ACP que permite realizar la reconstitución de la matriz con los datos faltantes estimados. Gonzales (2014).

La siguiente figura es el esquema básico del algoritmo NIPALS

El pseudocódigo de este algoritmo propuesto por Gonzales (2014), es el siguiente:

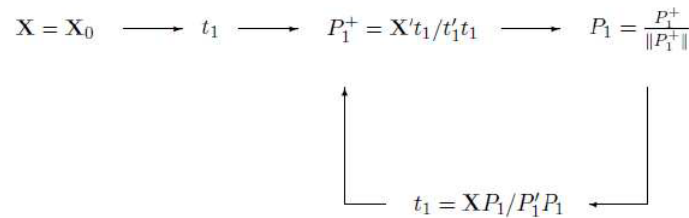


FIGURA 1: Esquema del algoritmo NIPALS

Etapla 1. $X_0 = X_h$

Etapla 2. $h = 1, 2, \dots, a$:

Etapla 2.1. $t_h = 1^a$ columna de X_{h-1}

Etapla 2.2. :Repetir hasta la convergencia de P_h

Etapla 2.2.1 Para $j = 1, 2, \dots, p$:

$$P_{hj} = \frac{\sum_{\{i: x_{ji} \text{ e } t_{hi} \text{ existen}\}} x_{h-1,ji} t_{hi}}{\sum_{\{i: x_{ji} \text{ e } t_{hi} \text{ existen}\}} t_{hi}^2} \quad [cov(t_h, x_{h-1,j})/s_{th}^2]$$

Etapla 2.2.2 normar P_h a 1.

Etapla 2.2.3 Para $i = 1, 2, \dots, n$: $t_{hi} = \frac{\sum_{\{j: x_{ji} \text{ existe}\}} x_{h-1,ji} P_{hj}}{\sum_{\{j: x_{ji} \text{ existe}\}} P_{hj}^2}$

Etapla 2.3. $X_h = X_{h-1} - t_h P_h'$

Fin

Para realizar este algoritmo NIPALS en matrices con datos faltantes, generamos NA's de forma aleatoria (se generaron pares de números correspondientes a las filas y columnas de la matriz y se reemplazaron por NA) en la matriz de datos correspondiente a la base de datos dada en la introducción , con 5,10,15, y 20 % de NA's o datos faltantes. Esas matrices son las siguientes:

- Matrices con el 5 % y 10 % de datos faltantes respectivamente:

Año	Colombia	Brasil	Chile	Argentina	Ecuador	Peru
1991	44.4	27.2	45.6	20.0	6.0	14.1
1992	75.5	11.8	58.9	22.6	NA	14.4
1993	110.7	50.6	128.3	17.2	119.4	118.5
1994	80.3	70.6	102.2	15.2	154.9	146.1
1995	81.6	82.3	89.0	35.1	169.4	127.1
1996	76.4	97.4	185.0	51.0	75.5	129.0
1997	32.0	89.5	195.3	31.1	33.4	NA
1998	55.5	63.1	66.3	24.4	9.7	66.7
1999	74.3	72.6	76.3	28.1	11.2	110.7
2000	84.5	76.2	80.1	29.5	11.8	110.2
2001	87.1	97.4	89.3	51.5	63.1	NA
2002	89.3	89.5	72.4	40.3	66.3	70.2
2003	70.2	63.1	80.1	60.5	76.3	90.1
2004	NA	66.3	70.5	39.1	20.0	64.5
2005	60.5	NA	107.2	31.1	63.4	92.7
2006	140.3	20.0	63.4	50.2	101.2	120.8
2007	120.4	22.6	101.2	51.0	103.1	107.2
2008	130.2	NA	103.1	42.5	66.7	70.8
2009	110.1	31.1	75.6	25.7	110.7	101.2
2010	120.2	24.4	68.9	60.3	110.2	110.8

Año	Colombia	Brasil	Chile	Argentina	Ecuador	Peru
1991	44.4	27.2	45.6	NA	6.0	14.1
1992	75.5	NA	58.9	22.6	17.8	14.4
1993	NA	50.6	NA	17.2	119.4	118.5
1994	80.3	70.6	102.2	15.2	154.9	146.1
1995	81.6	82.3	89.0	35.1	169.4	127.1
1996	76.4	97.4	185.0	51.0	75.5	129.0
1997	32.0	89.5	195.3	31.1	33.4	110.2
1998	55.5	63.1	66.3	24.4	9.7	66.7
1999	NA	72.6	NA	28.1	11.2	110.7
2000	84.5	76.2	80.1	29.5	11.8	110.2
2001	87.1	97.4	NA	51.5	63.1	89.3
2002	89.3	89.5	72.4	40.3	66.3	70.2
2003	70.2	63.1	80.1	60.5	76.3	90.1
2004	90.1	66.3	NA	39.1	20.0	64.5
2005	60.5	76.3	107.2	31.1	63.4	92.7
2006	140.3	20.0	63.4	50.2	101.2	120.8
2007	120.4	22.6	101.2	51.0	103.1	107.2
2008	NA	17.2	103.1	42.5	66.7	70.8
2009	NA	31.1	75.6	25.7	110.7	101.2
2010	120.2	24.4	68.9	NA	110.2	110.8

- Matrices con el 15 % y 20 % de datos faltantes respectivamente:

Año	Colombia	Brasil	Chile	Argentina	Ecuador	Peru	Año	Colombia	Brasil	Chile	Argentina	Ecuador	Peru
1991	NA	27.2	45.6	NA	6.0	14.1	1991	44.4	27.2	45.6	NA	6.0	14.1
1992	75.5	NA	58.9	22.6	17.8	14.4	1992	75.5	11.8	58.9	22.6	17.8	14.4
1993	110.7	50.6	128.3	NA	119.4	118.5	1993	NA	50.6	128.3	NA	119.4	118.5
1994	80.3	70.6	NA	15.2	154.9	146.1	1994	80.3	70.6	102.2	15.2	154.9	146.1
1995	81.6	82.3	89.0	35.1	169.4	NA	1995	81.6	82.3	89.0	35.1	169.4	NA
1996	76.4	97.4	185.0	51.0	75.5	NA	1996	76.4	97.4	NA	51.0	NA	129.0
1997	32.0	89.5	195.3	31.1	33.4	NA	1997	32.0	89.5	NA	NA	33.4	110.2
1998	55.5	63.1	NA	NA	9.7	66.7	1998	55.5	63.1	66.3	24.4	9.7	66.7
1999	74.3	72.6	76.3	28.1	11.2	NA	1999	74.3	72.6	NA	28.1	NA	110.7
2000	84.5	76.2	80.1	29.5	NA	110.2	2000	NA	76.2	80.1	29.5	11.8	110.2
2001	87.1	97.4	89.3	51.5	63.1	89.3	2001	87.1	NA	89.3	51.5	63.1	NA
2002	89.3	NA	72.4	40.3	66.3	70.2	2002	89.3	89.5	72.4	40.3	NA	NA
2003	70.2	63.1	80.1	60.5	76.3	90.1	2003	70.2	63.1	80.1	60.5	76.3	90.1
2004	90.1	66.3	NA	39.1	20.0	64.5	2004	90.1	66.3	70.5	39.1	20.0	NA
2005	60.5	76.3	NA	31.1	63.4	92.7	2005	60.5	NA	107.2	31.1	63.4	92.7
2006	140.3	20.0	63.4	50.2	101.2	120.8	2006	140.3	20.0	63.4	50.2	101.2	120.8
2007	120.4	22.6	101.2	51.0	103.1	107.2	2007	120.4	NA	NA	51.0	103.1	107.2
2008	NA	17.2	NA	42.5	66.7	70.8	2008	130.2	17.2	103.1	42.5	66.7	70.8
2009	110.1	31.1	NA	25.7	110.7	101.2	2009	110.1	31.1	75.6	25.7	NA	NA
2010	120.2	24.4	68.9	60.3	110.2	110.8	2010	120.2	24.4	68.9	60.3	NA	110.8

3.2. Resultados

Los valores propios obtenidos mediante el algoritmo NIPALS para las matrices con 5,10,15 y 20 % son:

% NA	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
5%	2,4795911	1,9574160	0,9418925	0,5887012	0,2926197	0,1261818
10%	2,156179	1,9495680	1,1627291	0,4743410	0,2712866	0,1075340
15%	2,7541533	2,0189161	0,8745523	0,3612268	0,1725610	0,1452363
20%	3,01079935	1,8538082	0,7589054	0,4764867	0,2343548	0,0456656

En cuanto a los vectores propios obtenidos, por cuestiones de espacio, compararemos solo los tres primeros vectores propios de cada matriz:

	5%			10%			15%			20%		
	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3
Colombia	0,429714	0,462774	0,010940	0,489972	0,337622	0,239236	0,541172	0,226588	0,010857	0,453184	0,437367	0,271795
Brasil	0,072810	-0,630665	0,172080	-0,222065	-0,624667	0,087050	-0,376927	-0,449971	-0,224875	0,031260	-0,680329	-0,301077
Chile	0,190536	-0,557458	0,061648	0,055773	-0,577781	0,189901	-0,179969	-0,603601	-0,275110	0,398640	-0,291447	0,157374
Argentina	0,307943	0,120163	0,902719	0,235302	-0,010222	0,885203	0,253161	0,174995	-0,919974	0,589108	0,215449	-0,677587
Ecuador	0,614783	0,015342	-0,359984	0,603334	-0,047589	-0,299120	0,452037	-0,391397	0,164834	0,394665	-0,216978	0,581645
Perú	0,548592	-0,250335	-0,148247	0,536774	-0,399503	-0,161459	0,514044	-0,445012	-0,010266	0,363215	-0,409174	0,115233

Las componentes principales, también mostraremos las primeras tres para cada matriz:

	5%			10%			15%			20%		
	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3
1991	-3,38561	0,96224	-0,52252	-2,66750	2,04076	-1,39643	-1,65576	2,69941	0,92977	-4,05012	1,08850	0,73527
1992	-3,17064	1,63906	-0,25508	-2,41899	1,85653	-0,35675	-2,32608	1,57941	1,11585	-2,52801	1,97281	0,42393
1993	1,10355	-0,23717	-1,70741	0,99864	-0,02049	-1,85360	1,49240	-0,93401	-0,64793	3,46691	0,34316	-0,07606
1994	1,39985	-0,98069	-2,15301	1,42211	-1,11806	-2,24315	0,91680	-3,08385	1,69149	0,48462	-2,03815	2,19461
1995	1,72311	-0,69924	-0,82566	1,58126	-0,99349	-0,87997	0,58190	-1,70723	0,54383	0,85547	-1,67174	1,09327
1996	1,34960	-2,47212	1,17382	0,64130	-2,69809	1,27803	-1,19477	-1,92428	-1,67327	1,15153	-1,75298	-0,85475
1997	-1,20286	-3,25403	0,28139	-1,27443	-2,88730	-0,02913	-3,45986	-1,57594	-0,48362	-1,85495	-2,04811	-0,36828
1998	-2,06252	-0,24303	-0,22459	-1,92016	0,41734	-0,54820	-1,79814	0,35036	0,07033	-2,11121	-0,19723	-0,39885
1999	-0,92232	-0,52498	-0,09807	-0,93077	-0,96685	-0,30734	-1,60896	0,87904	0,49460	-0,67934	-0,95321	0,41531
2000	-0,70789	-0,47074	0,02278	-0,67744	-0,27090	-0,12182	-0,09366	-0,60951	0,69882	-0,85860	-0,96250	-0,60835
2001	0,44962	-0,76668	1,26155	-0,06534	-1,13176	1,47463	-0,32314	-0,34161	-1,19233	0,82211	0,09686	-0,69267
2002	-0,31264	-0,16603	0,55745	-0,45790	-0,00560	0,52067	-0,14854	0,75606	0,07256	-0,00793	-0,77645	-0,58155
2003	0,25757	0,06666	1,56848	0,23652	-0,02401	1,51578	0,14900	0,19297	-1,46217	0,73316	0,00588	-1,29158
2004	-1,39243	0,03556	0,70259	-0,92082	0,59051	0,76826	-0,89693	0,86773	-0,22229	-0,55006	0,28600	-0,72060
2005	-0,55151	-1,13710	-0,14717	-0,67308	-0,80694	-0,34001	-0,99591	-1,12045	0,18173	-0,27316	-1,57939	-0,08401
2006	1,68383	2,19068	0,29995	2,43904	1,64462	0,81415	2,80179	0,92929	-0,13159	1,50502	1,80212	0,65461
2007	1,41452	1,33701	0,46841	1,93462	0,96626	0,94906	2,01778	0,37995	-0,44040	1,78305	0,61257	0,31577
2008	0,44845	1,27030	0,44834	0,12543	1,26578	0,65730	0,51803	1,85137	0,37723	1,02225	1,69975	0,76010
2009	0,58818	1,14563	-1,21805	0,87570	1,02993	-1,08417	1,23473	0,10498	1,53413	-0,45739	1,20662	1,33149
2010	1,60743	1,84073	0,96754	1,84051	1,34595	0,32946	2,42527	0,80440	-0,89674	1,64705	1,66759	-0,47150

3.3. EM-PCA

El método Análisis de Componentes Principales vía EM (ACP-EM), es un método propuesto por Josse and Husson (2012), el cual realiza un Análisis de Componentes principales en presencia de datos faltantes. Particularmente, los datos faltantes son estimados inicialmente por valores promedio y luego se desea minimizar la distancia entre los valores originales (Z) y los valores estimados ($\Psi u'$). Ochoa (2018)

Para realizar este algoritmo en R Core Team (2017), usamos la función del paquete ade4. El pseudocódigo de este algoritmo propuesto por Ochoa (2018) es:

1. Iniciación $L=0$
 Z^0 los datos faltantes son reemplazados por valores iniciales, como por ejemplo la media.
2. Paso L
 - 2.1 Realice un ACP para calcular Ψ^L, u^L , tomando q dimensiones (para garantizar un proceso iterativo).
 - 2.2 Los valores faltantes son imputados via reconstitución de la matriz $Z^L = \Psi^L u'^L$. Los valores observados son los mismos y los faltantes se reemplazan por la imputación.
3. Se repiten el paso 2.1 y 2.2 hasta la convergencia.

3.4. Resultados

Los valores propios obtenidos mediante el algoritmo EM-PCA para las matrices con 5,10,15 y 20 % de datos faltantes son:

% NA	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
5%	2,235339	1,923973	0,904790	0,481629	0,288574	0,165697
10%	2,106891	1,988887	1,044122	0,429413	0,295421	0,135266
15%	2,422113	2,142823	0,850670	0,360214	0,149434	0,074747
20%	2,712656	1,838548	0,669018	0,460891	0,226887	0,092000

En cuanto a los vectores propios obtenidos, por cuestiones de espacio, compararemos solo los tres primeros vectores propios de cada matriz:

	5%			10%			15%			20%		
	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3
Colombia	-0,334085	0,562250	0,016244	0,590850	0,241440	0,141372	0,601545	-0,114905	-0,082281	-0,268283	0,614693	0,189306
Brasil	-0,190819	-0,602094	0,192583	-0,247043	-0,558790	0,062836	-0,552897	-0,156642	0,207208	-0,224581	-0,627557	-0,378188
Chile	-0,338732	-0,474736	0,001897	-0,044784	-0,621902	0,162562	-0,423327	-0,403031	0,167667	-0,460454	-0,161048	0,355306
Argentina	-0,297174	0,190883	0,884138	0,248249	-0,052183	0,892656	0,324347	-0,102589	0,916765	-0,291383	0,412317	-0,784264
Ecuador	-0,530656	0,206737	-0,394717	0,569424	-0,151401	-0,328128	0,214939	-0,587164	-0,284671	-0,532033	0,046700	0,270350
Perú	-0,606076	-0,129639	-0,158563	0,449432	-0,465889	-0,212453	0,042927	-0,666739	-0,026216	-0,545534	-0,173788	-0,082067

Las componentes principales, también mostraremos las primeras tres para cada matriz:

	5%			10%			15%			20%		
	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3
1991	3,626683	0,263291	-0,407144	-2,489497	2,449398	-0,072293	0,905144	3,023943	-0,016483	3,909336	0,095974	-0,020289
1992	3,032592	1,160975	-0,442228	-1,927845	2,120317	-0,348764	-0,663319	2,767307	-0,977592	3,343545	0,949753	0,920148
1993	-1,132140	0,086053	-1,795388	0,810795	-0,210146	-1,818897	0,724626	-1,600365	-0,008707	-2,434962	0,450429	0,736078
1994	-1,512373	-0,625419	-2,220552	1,117358	-1,296318	-2,369615	-1,163878	-2,502403	-2,024493	-1,855701	-1,515407	1,851160
1995	-1,745381	-0,282198	-0,885654	1,381125	-1,174559	-0,919216	-0,215698	-1,921839	-0,729780	-1,990789	-0,929030	0,371490
1996	-1,997177	-2,132116	1,012304	0,235342	-2,977362	1,306279	-1,723843	-1,869548	1,611287	-1,986335	-1,057800	-0,981527
1997	-0,240878	-3,364454	0,089671	-1,910353	-2,942963	0,031558	-3,516666	-0,476277	0,508840	0,666910	-2,492531	-0,338078
1998	1,980866	-0,700421	-0,134152	-1,940659	0,738460	-0,528460	-1,407104	1,397469	0,112845	2,111362	-1,015715	-0,105872
1999	0,718712	-0,768679	-0,033159	-1,104702	-0,651665	-0,268769	-1,055183	1,677667	-0,344223	-0,196218	-1,057039	0,319748
2000	0,508175	-0,667093	0,086340	-0,686965	-0,114670	-0,184516	-0,524893	-0,289411	-0,658343	0,741817	-1,354116	-0,385519
2001	-0,765054	-0,670775	1,244572	-0,082201	-1,047051	1,410679	-0,507684	-0,153263	1,302952	-0,422449	0,478298	-0,688538
2002	0,299760	-0,232956	0,617262	-0,350216	0,114220	0,506996	0,436070	0,731504	-0,014041	-0,291849	-0,579170	-0,794638
2003	-0,254661	0,122275	1,544169	0,223698	-0,038841	1,654370	0,223500	-0,029584	1,644355	-0,330933	0,272925	-1,572314
2004	1,294978	-0,177448	0,772776	-0,780737	0,609152	0,780615	-0,280835	1,211849	0,365183	0,992577	0,037467	-0,690897
2005	0,306052	-1,030583	-0,215734	-0,878562	-0,741560	-0,325897	-1,569001	-0,105523	-0,205091	-0,247342	-1,421963	0,378993
2006	-1,238409	2,563463	0,264703	2,800444	1,375903	0,761932	2,927646	-0,756923	0,029523	-0,779641	2,537846	-0,097946
2007	-1,156070	1,694894	0,372717	2,116294	0,660415	0,953951	1,987492	-0,832517	0,339302	-1,256976	1,364915	-0,119585
2008	-0,216583	1,230810	0,460000	0,404850	1,115499	0,643000	1,798624	0,939959	-0,232576	-0,246886	2,023793	1,008219
2009	-0,290754	1,324339	-1,238356	1,024160	0,863840	-1,133275	1,043013	-0,462783	-1,578681	0,947230	0,911045	1,060168
2010	-1,218338	2,206041	0,907852	2,037671	1,147930	-0,079680	2,581990	-0,749264	0,875722	-0,672694	2,300328	-0,850800

4. Comparaciones

1. Poder descriptivo en los dos primeros ejes

El poder descriptivo de las dos primeras componentes para los tres métodos y las matrices generadas con el 5,10,15 y 20% de datos faltantes se resume en la siguiente tabla:

	ACP completo	Nipals				ACP-EM			
		5%	10%	15%	20%	5%	10%	15%	20%
Poder descriptivo	0,691598	0,694759	0,670694	0,754439	0,762475	0,693219	0,682630	0,760823	0,758534

2. Valores propios

La comparación de los valores propios se puede ver en la siguiente tabla:

	%NA	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
ACP	Completa	2,193409	1,956178	0,903879	0,511947	0,285441	0,149146
Nipals	5%	2,479591	1,957416	0,941893	0,588701	0,292620	0,126182
	10%	2,156179	1,949568	1,162729	0,474341	0,271287	0,107534
	15%	2,754153	2,018916	0,874552	0,361227	0,172561	0,145236
	20%	3,010799	1,853808	0,758905	0,476487	0,234355	0,045666
ACP-EM	5%	2,235339	1,923973	0,904790	0,481629	0,288574	0,165697
	10%	2,106891	1,988887	1,044122	0,429413	0,295421	0,135266
	15%	2,422113	2,142823	0,850670	0,360214	0,149434	0,074747
	20%	2,712656	1,838548	0,669018	0,460891	0,226887	0,092000

3. Vectores propios

La comparación de los dos primeros vectores propios se puede ver en la siguiente tabla:

ACP		NIPALS								ACP-EM							
Completa		5%		10%		15%		20%		5%		10%		15%		20%	
λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2
-0,3267	0,5709	0,4297	0,4628	0,4900	0,3376	0,5412	0,2266	0,4532	0,4374	-0,3341	0,5622	0,5909	0,2414	0,6015	-0,1149	-0,2683	0,6147
-0,1589	-0,6033	0,0728	-0,6307	-0,2221	-0,6247	-0,3769	-0,4500	0,0313	-0,6803	-0,1908	-0,6021	-0,2470	-0,5588	-0,5529	-0,1566	-0,2246	-0,6276
-0,3382	-0,4636	0,1905	-0,5575	0,0558	-0,5778	-0,1800	-0,6036	0,3986	-0,2914	-0,3387	-0,4747	-0,0448	-0,6219	-0,4233	-0,4030	-0,4605	-0,1610
-0,2899	0,2066	0,3079	0,1202	0,2353	-0,0102	0,2532	0,1750	0,5891	0,2154	-0,2972	0,1909	0,2482	-0,0522	0,3243	-0,1026	-0,2914	0,4123
-0,5459	0,1756	0,6148	0,0153	0,6033	-0,0476	0,4520	-0,3914	0,3947	-0,2170	-0,5307	0,2067	0,5694	-0,1514	0,2149	-0,5872	-0,5320	0,0467
-0,6096	-0,1470	0,5486	-0,2503	0,5368	-0,3995	0,5140	-0,4450	0,3632	-0,4092	-0,6061	-0,1296	0,4494	-0,4659	0,0429	-0,6667	-0,5455	-0,1738

4. Componentes principales

La comparación de las dos primeras componentes principales se puede ver en la siguiente tabla:

	ACP		NIPALS								ACP-EM							
	Completa		5%		10%		15%		20%		5%		10%		15%		20%	
	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2
1991	-3,5794	0,2200	-3,3856	0,9622	-2,6675	2,0408	-1,6558	2,6994	-4,0501	1,0885	3,6267	0,2633	-2,4895	2,4494	0,9051	3,0239	3,9093	0,0960
1992	-2,9842	1,1042	-3,1706	1,6391	-2,4190	1,8565	-2,3261	1,5794	-2,5280	1,9728	3,0326	1,1610	-1,9278	2,1203	-0,6633	2,7673	3,3435	0,9498
1993	1,1801	-0,0116	1,1035	-0,2372	0,9986	-0,0205	1,4924	-0,9340	3,4669	0,3432	-1,1321	0,0861	0,8108	-0,2101	0,7246	-1,6004	-2,4350	0,4504
1994	1,5534	-0,7594	1,3998	-0,9807	1,4221	-1,1181	0,9168	-3,0839	0,4846	-2,0382	-1,5124	-0,6254	1,1174	-1,2963	-1,1639	-2,5024	-1,8557	-1,5154
1995	1,7594	-0,3827	1,7231	-0,6992	1,5813	-0,9935	0,5819	-1,7072	0,8555	-1,6717	-1,7454	-0,2822	1,3811	-1,1746	-0,2157	-1,9218	-1,9908	-0,9290
1996	1,9519	-2,1097	1,3496	-2,4721	0,6413	-2,6981	-1,1948	-1,9243	1,1515	-1,7530	-1,9972	-2,1321	0,2353	-2,9774	-1,7238	-1,8695	-1,9863	-1,0578
1997	-0,2414	-3,3571	-1,2029	-3,2540	-1,2744	-2,8873	-3,4599	-1,5759	-1,8549	-2,0481	-0,2409	-3,3645	-1,9104	-2,9430	-3,5167	-0,4763	0,6669	-2,4925
1998	-1,9834	-0,7162	-2,0625	-0,2430	-1,9202	0,4173	-1,7981	0,3504	-2,1112	-0,1972	1,9809	-0,7004	-1,9407	0,7385	-1,4071	1,3975	2,1114	-1,0157
1999	-0,7344	-0,7829	-0,9223	-0,5250	-0,9308	-0,9669	-1,6090	0,8790	-0,6793	-0,9532	0,7187	-0,7687	-1,1047	-0,6517	-1,0552	1,6777	-0,1962	-1,0570
2000	-0,5323	-0,6705	-0,7079	-0,4707	-0,6774	-0,2709	-0,0937	-0,6095	-0,8586	-0,9625	0,5082	-0,6671	-0,6870	-0,1147	-0,5249	-0,2894	0,7418	-1,3541
2001	0,3726	-0,5713	0,4496	-0,7667	-0,0653	-1,1318	-0,3231	-0,3416	0,8221	0,0969	-0,7651	-0,6708	-0,0822	-1,0471	-0,5077	-0,1533	-0,4224	0,4783
2002	-0,3406	-0,2237	-0,3126	-0,1660	-0,4579	-0,0056	-0,1485	0,7561	-0,0079	-0,7765	0,2998	-0,2330	-0,3502	0,1142	0,4361	0,7315	-0,2918	-0,5792
2003	0,2531	0,1018	0,2576	0,0667	0,2365	-0,0240	0,1490	0,1930	0,7332	0,0059	-0,2547	0,1223	0,2237	-0,0388	0,2235	-0,0296	-0,3309	0,2729
2004	-1,1314	0,1457	-1,3924	0,0356	-0,9208	0,5905	-0,8969	0,8677	-0,5501	0,2860	1,2950	-0,1774	-0,7807	0,6092	-0,2808	1,2118	0,9926	0,0375
2005	-0,2640	-1,2144	-0,5515	-1,1371	-0,6731	-0,8069	-0,9959	-1,1204	-0,2732	-1,5794	0,3061	-1,0306	-0,8786	-0,7416	-1,5690	-0,1055	-0,2473	-1,4220
2006	1,3042	2,4709	1,6838	2,1907	2,4390	1,6446	2,8018	0,9293	1,5050	1,8021	-1,2384	2,5635	2,8004	1,3759	2,9276	-0,7569	-0,7796	2,5378
2007	1,2198	1,6157	1,4145	1,3370	1,9346	0,9663	2,0178	0,3800	1,7830	0,6126	-1,1561	1,6949	2,1163	0,6604	1,9875	-0,8325	-1,2570	1,3649
2008	0,0769	1,8051	0,4484	1,2703	0,1254	1,2658	0,5180	1,8514	1,0222	1,6998	-0,2166	1,2308	0,4048	1,1155	1,7986	0,9400	-0,2469	2,0238
2009	0,3592	1,2125	0,5882	1,1456	0,8757	1,0299	1,2347	0,1050	-0,4574	1,2066	-0,2908	1,3243	1,0242	0,8638	1,0430	-0,4628	0,9472	0,9110
2010	1,2776	2,1234	1,6074	1,8407	1,8405	1,3459	2,4253	0,8044	1,6471	1,6676	-1,2183	2,2060	2,0377	1,1479	2,5820	-0,7493	-0,6727	2,3003

5. Correlación entre la componente 1 completo y faltantes

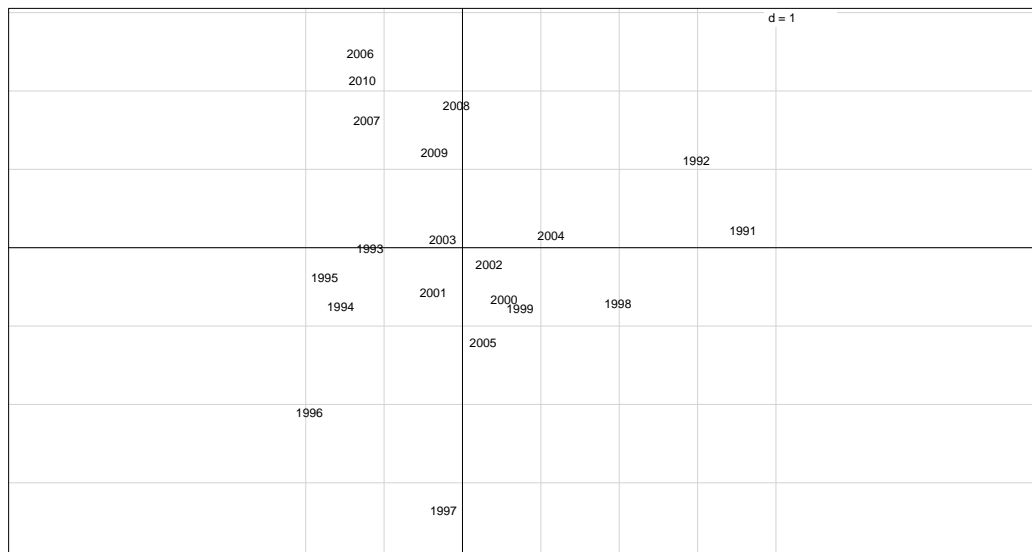
Para comparar estas correlaciones, se obtendrá el valor absoluto de la correlación entre la componente 1 de la matriz completa y la componente 1 de cada una de las matrices incompletas por NIPLALS y por ACP-EM. Se sacará el valor absoluto ya que por cuestiones del código, los signos entre una componente y otra pueden ser iguales o inversos.

	Nipals				ACP-EM			
	5%	10%	15%	20%	5%	10%	15%	20%
Correlación Comp 1	0,966200	0,904884	0,620512	0,851966	0,997426	0,821209	0,163642	0,936184

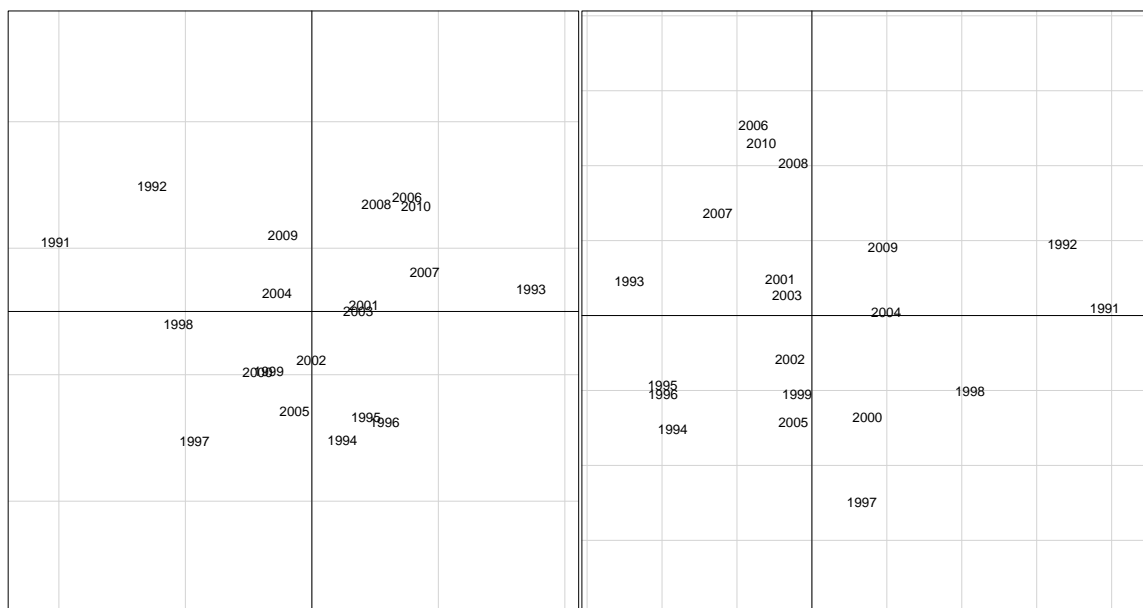
6. Estructura factorial en la nube de individuos

Para comparar estas estructuras, solo se comparará el ACP con datos completos con el ACP-EM y el NIPALS para la matriz con más datos faltantes (20 % de datos faltantes)

La nube de individuos correspondiente al ACP de la matriz de datos completos es



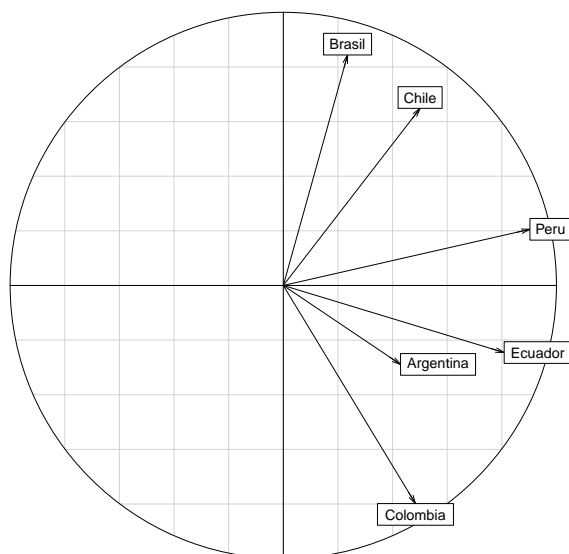
Las nubes de individuos para el NIPALS y el ACP-EM con una matriz del 20 % de datos faltantes es respectivamente



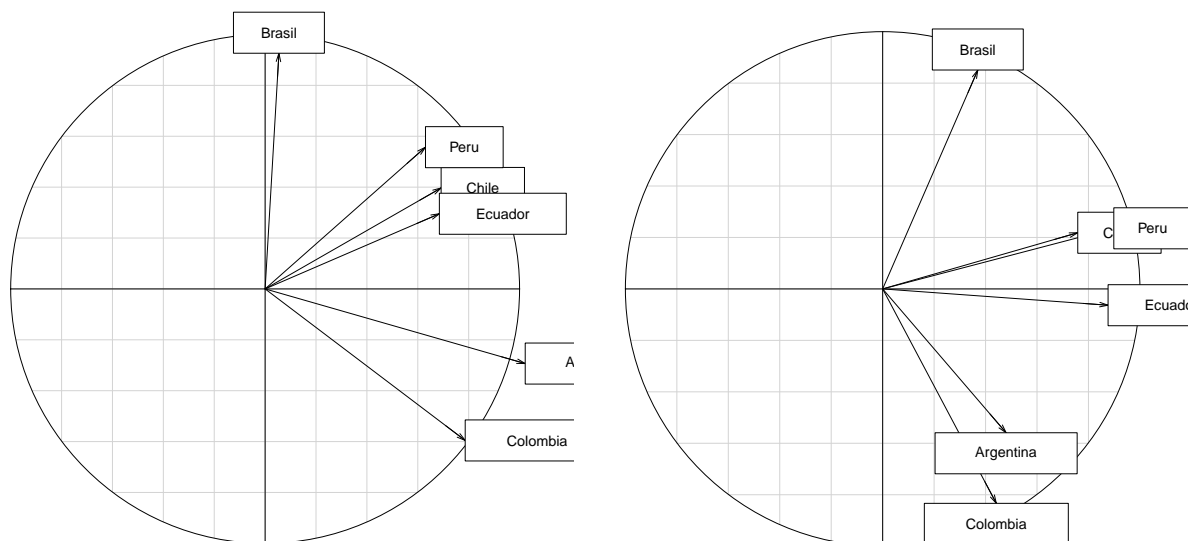
7. Estructura factorial en la nube de variables

Para comparar estas estructuras, solo se comparará el ACP con datos completos con el ACP-EM y el NIPALS para la matriz con más datos faltantes (20 % de datos faltantes)

La nube de variables correspondiente al ACP de la matriz de datos completos es



Las nubes de variables para el NIPALS y el ACP-EM con una matriz del 20 % de datos faltantes es respectivamente



8. Ortogonalidad en las dos primeras componentes y ortonormalidad en los dos primeros vectores propios

La ortogonalidad en las dos primeras componentes se puede evidenciar en la siguiente tabla, donde se hizo el producto punto entre la componente 1 y la componente 2

	ACP completo	Nipals				ACP-EM			
		5%	10%	15%	20%	5%	10%	15%	20%
Ortogonalidad	0	0,686668	-1,395858	-0,819374	1,147813	0	0	0	0

La ortonormalidad en los dos primeros vectores propios se puede ver en la siguiente tabla, donde se obtuvo la norma de cada uno, y posteriormente se hizo su producto punto.

	ACP		NIPALS								ACP-EM							
	Completa		5%		10%		15%		20%		5%		10%		15%		20%	
	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2
Norma	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ortogonalidad	0		-0,044171		0,026357		0,039479		-0,046576		0		0		0		0	

9. Imputación

Referencias

- Gonzales, V. M. (2014), ANÁLISIS CONJUNTO DE MÚLTIPLES TABLAS DE DATOS MIXTOS MEDIANTE PLS, PhD thesis, UNIVERSITAT POLITÈCNICA DE CATALUNYA, Barcelona.
- Kassambara, A. & Mundt, F. (2017), *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. R package version 1.0.5.
*<https://CRAN.R-project.org/package=factoextra>
- Lê, S., Josse, J. & Husson, F. (2008), ‘FactoMineR: A package for multivariate analysis’, *Journal of Statistical Software* **25**(1), 1–18.
- Ludovic Lebart, Alain Morineau, M. P. (1995), *Statistique exploratoire multidimensionnelle*, Dunod, Paris.
- Ochoa, A. F. (2018), Análisis de correspondencias múltiples en presencia de datos faltantes: el principio de datos disponibles del algoritmo nipals (acmpdd), Master’s thesis, Universidad del Valle.
- R Core Team (2017), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
*<https://www.R-project.org/>
- Wickham, H. (2009), *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag New York.
*<http://ggplot2.org>
- Wickham, H. & Bryan, J. (2018), *readxl: Read Excel Files*. R package version 1.1.0.
*<https://CRAN.R-project.org/package=readxl>
- Zelaya, J. T. (n.d.), *ANÁLISIS MULTIVARIADO DE DATOS*, Universidad de Costa Rica.