

1ª Práctica de ACP

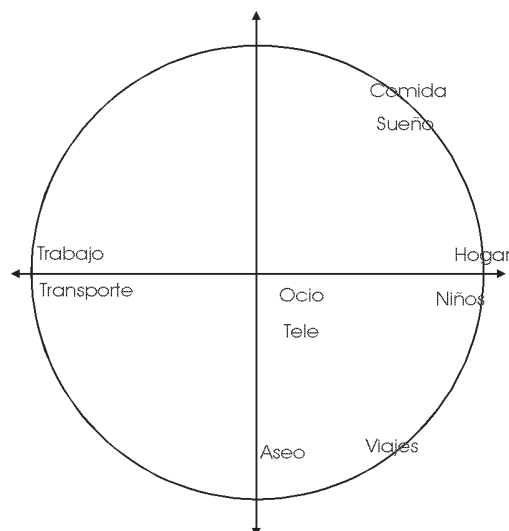
La tabla siguiente correspondiente al ejemplo nº 3 del primer tema “*Estudio de diversas tablas de datos*” se compone de 28 clases de sujetos, cuyas variables de base son: el sexo, el país, la actividad profesional y el estado civil, y por otro lado 10 clases de actividades reagrupadas por la necesidad de comparar el tiempo de dedicación a ellas. Un caso de la tabla, contiene el número de horas que los sujetos de la clase i han dedicado por término medio a la actividad j , durante la duración de la encuesta.

nº horas	Prof	Tran	Mena	Enfa	Cour	Toil	Repa	Somm	Tele	Lois
HAUS	610	140	60	10	120	95	115	760	175	315
FAUS	475	90	250	30	140	120	100	775	115	305
FNAU	10	0	495	110	170	110	130	785	160	430
HMUS	615	140	65	10	115	90	115	765	180	305
FMUS	179	29	421	87	161	112	119	776	143	373
HCUS	585	115	50	0	150	105	100	760	150	385
FCUS	482	94	196	18	141	130	96	775	132	336
HAWE	653	100	95	7	57	85	150	808	115	330
FAWE	511	70	307	30	80	95	142	816	87	262
FNAW	20	7	568	87	112	90	180	843	125	368
HMWE	656	97	97	10	52	85	152	808	122	321
FMWE	168	22	528	69	102	83	174	824	119	311
HCWE	643	105	72	0	62	77	140	813	100	388
FCWE	429	34	262	14	92	97	147	849	84	392
HAYO	650	140	120	15	85	90	105	760	70	365
FAYO	560	105	375	45	90	90	95	745	60	235
FNAY	10	10	710	55	145	85	130	815	60	380
HMYO	650	145	112	15	85	90	105	760	80	358
FMYO	260	52	576	59	116	85	117	775	65	295
HCYO	615	125	95	0	115	90	85	760	40	475
FCYO	433	89	318	23	112	96	102	774	45	408
HAES	650	142	122	22	76	94	100	764	96	334
FAES	578	106	338	42	106	94	92	752	64	228
FNAE	24	8	594	72	158	92	128	840	86	398
HMES	652	133	134	22	68	94	102	763	122	310
FMES	436	79	433	60	119	90	107	772	73	231
HCES	627	148	68	0	88	92	86	770	58	463
FCES	434	86	297	21	129	102	94	799	58	380

Interpretación y Conclusiones

1. Análisis de la nube de variables:

Variables en un círculo de centro 0 y radio 1



- Según se puede observar en la representación gráfica de las variables sobre un círculo de centro 0 y radio 1, hay dos grupos de variables, las formadas por Hogar y Niños y por Trabajo y Transporte que tienen unos valores de coordenadas respecto al primer eje, próximos a 1 y -1 respectivamente, que nos indican por un lado que hay una correlación cercana a -1 entre ambos grupos de variables, es decir que si un grupo de individuos aumenta su valor sobre uno de los grupos, ese mismo grupo de individuos lo disminuye sobre el otro, y por otro lado al tener coordenadas próximas a 1 en valor absoluto, establecen para los $F_{\alpha}(i)$, una clasificación de los individuos sobre este primer eje según el orden de los valores crecientes y decrecientes respectivamente sobre esos dos grupos de variables, o dicho de otra forma, que las coordenadas de los individuos para esos dos grupos de variables, son proporcionales directa e inversa respectivamente, a la desviación entre la media de esas variables y el valor de los individuos en esas variables.

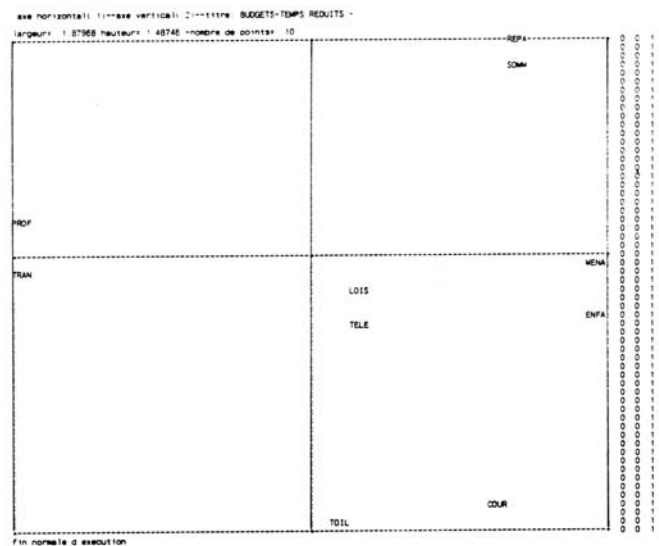
- Dentro de cada grupo de variables, la correlación existente entre las variables que los constituyen está próxima a uno, lo que hace que la distancia entre esas variables esté cercana a cero: $d^2(j, j') = 2(1 - \rho_{jj'})$. Los grupos de individuos que tienen una alta dedicación a tareas del hogar también lo tienen a la atención a los niños, y los que dedican mayor tiempo al trabajo también lo hacen con el transporte.
- Las variables Ocio, Tele y Aseo presentan un valor de coordenada respecto al primer eje, cercano al cero, por lo que no presentan relación lineal con los valores que toman los individuos sobre esas variables. Asimismo no están correlacionadas con los dos anteriores grupos de variables.
- Las variables Comida y Sueño muy correlacionadas entre sí, y la variable Viajes, al estar muy cerca del borde del círculo están muy bien representadas sobre el plano formado por los dos ejes, aunque no lo estén en cada uno de ellos por separado. Presentan también una correlación alta, aunque no cercana a uno, con el grupo de variables Hogar y Niños, porque gráficamente se puede observar el coeficiente de correlación, por la proyección ortogonal de las variables sobre el primer eje, al estar Hogar y Niños sobre dicho eje.
- La variable Aseo, al tener un valor de coordenada cercano a -1 sobre el 2º eje, establece también una clasificación de los individuos según valores decrecientes en relación a los valores sobre esa variable, aunque no tan fuerte como la establecida por las variables, Hogar-Niños y Trabajo-Transporte sobre el 1º eje.
- El 3º eje establece una clara separación entre la variable Televisión y el resto de variables, y como el valor de su coordenada está muy próxima a -1, (-0.93), establece asimismo una clasificación de los individuos en el orden de los valores decrecientes de esa variable.
- El 4º eje establece la separación entre el tiempo dedicado al Ocio y el resto de variables, ordenando a los individuos también en sentido decreciente de sus valores en esta variable.

2. Análisis de la nube de individuos:

- En relación al primer eje encontramos por lo tanto que los grupos: FNAU, (mujeres no activas de USA), FNAE, (mujeres no activas del Este), FNAW, (mujeres no activas del Oeste), FNAY, (mujeres no activas de Yugoslavia), FMWE, (mujeres casadas del Oeste) y FMUS, (mujeres casadas de USA), son los grupos de individuos que más tiempo dedican a la atención a las tareas del hogar y a la atención a los niños, por la clasificación que estas variables establecen para los grupos de individuos, y los que menos dedican a las variables trabajo y transporte.
- Los grupos de individuos: HCES (hombres solteros del Este), HAYO (hombres activos de Yugoslavia), HMYO (hombres casados de Yugoslavia), HMES (hombres casados del Este), HAES (hombres activos del Este), HCYO (hombres solteros de Yugoslavia), HMUS (hombres casados de USA), HAUS (hombres activos de USA) y HCUS (hombres solteros de USA), presentan una alta dedicación al trabajo y transporte relacionado con el trabajo y muy poco al hogar y los niños.
- Los grupos de individuos: HCWE (hombres solteros del Oeste), HAWO (hombres activos del Oeste), HMWE (hombres casados del Oeste), y por otro lado, FAYO (mujeres activas de Yugoslavia) y FAES (mujeres activas del Este), presentan también valores altos en relación al tiempo dedicado al trabajo y al transporte, aunque menos que los anteriores grupos observando el primer eje. En cambio en relación al segundo eje los hombres presentan valores más alejados que las mujeres en el tiempo invertido en aseo y viajes.
- Hay varios grupos de individuos: FCUS y FAUS, con valores altos en Aseo y Viajes, FAWO y FCWE, con valores altos en comida y sueño, FMYO, con valores altos en hogar y niños y FCES, FCYO y FMES, que presentan valores muy próximos a los valores medios de las variables hogar-niños y trabajo-transporte.
- Observando el segundo eje encontramos que los grupos de individuos: HCWE, HMWE, HAWO, FMWE, FNAW, FAWO, (mujeres activas del Oeste), FCWE, (mujeres solteras del Oeste), dedican mayor tiempo a la comida y al sueño y menos al aseo personal y viajes, en contra de lo que ocurre con los grupos de individuos: HCUS, FAUS, (mujeres activas de USA), FCUS, (mujeres solteras de USA), FMUS y FNAU que invierten más tiempo al aseo personal y viajes y menos a la comida y a dormir.
- Hay un grupo muy numeroso de individuos que mantienen valores cercanos a la media en relación a las variables comida-sueño y aseo-viajes, como son: HAYO, HMES, HAES, HCES, HCYO, HAUS, HMUS, FAYO, FAES, FCYO, FCES, FMES, FMYO, FNAY y FNAE, (2º eje).

- ### 3. Significado de los ejes:

- figura 1: Individuos y variables



Análisis de componentes principales con MVSP. Práctica 1

Pasos a realizar:

1. Crea la carpeta ACP.
2. Crea dentro de ACP con el editor 'Bloc de notas' el fichero TIEMPOS.MVS, cuyo contenido debe ser:

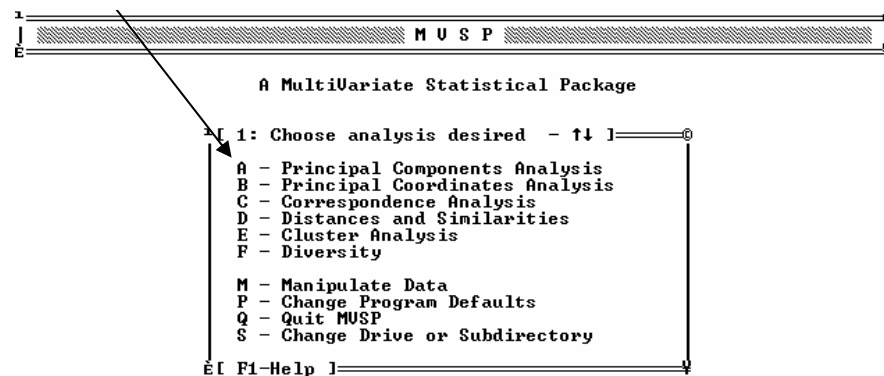
```

Prof Tran Mena Enfa Cour Toil Repa Somm Tele Lois
HAUS 610 140 60 10 120 95 115 760 175 315
FAUS 475 90 250 30 140 120 100 775 115 305
FNAU 10 0 495 110 170 110 130 785 160 430
HMUS 615 140 65 10 115 90 115 765 180 305
FMUS 179 29 421 87 161 112 119 776 143 373
HCUS 585 115 50 0 150 105 100 760 150 385
FCUS 482 94 196 18 141 130 96 775 132 336
HAME 653 100 95 7 57 85 150 808 115 330
FAWE 511 70 307 30 80 95 142 816 87 262
FNAW 20 7 568 87 112 90 180 843 125 368
HMWE 656 97 97 10 52 85 152 808 122 321
FMWE 168 22 528 69 102 83 174 824 119 311
HCWE 643 105 72 0 62 77 140 813 100 388
FCWE 429 34 262 14 92 97 147 849 84 392
HAYO 650 140 120 15 85 90 105 760 70 365
FAYO 560 105 375 45 90 90 95 745 60 235
FNAY 10 10 710 55 145 85 130 815 60 380
HMYO 650 145 112 15 85 90 105 760 80 358
FMYO 260 52 576 59 116 85 117 775 65 295
HCYO 615 125 95 0 115 90 85 760 40 475
FCYO 433 89 318 23 112 96 102 774 45 408
HAES 650 142 122 22 76 94 100 764 96 334
FAES 578 106 338 42 106 94 92 752 64 228
FNAE 24 8 594 72 158 92 128 840 86 398
HMES 652 133 134 22 68 94 102 763 122 310
FMES 436 79 433 60 119 90 107 772 73 231
HCES 627 148 68 0 88 92 86 770 58 463
FCES 434 86 297 21 129 102 94 799 58 380

```

3. Ejecuta el fichero C:\MVSP\MVSPSHARE.EXE
4. Elige la opción A , e introduce el nombre del fichero de datos: tiempos.mvs

Enter name of input data file (? for directory): a:\ACP\tiempos.mvs



Active drive\path - D:\STATIS\MUSP

4. Proporciona la información pedida por el programa:

```

Input file - A:\ACP\TIEMPOS.MUS
Data file header not found.
The following is the first line of the file:

Prof Tran Mena Enfa Cour Toil Repa Somm Tele Lois

Is this the file you wish to analyze? Y
Please provide the following information:

Are data labels included in the data file? Y
Enter the number of rows
=> 028
Enter the number of columns
=> 010

```

5. Marca las siguientes tres opciones:

T - Transpose Data	Default: Yes
C - Center Data	Default: Yes
D - Standardize Data	Default: Yes
6. Marcar la opción RUN R - Run Analysis
7. Introduce el nombre del fichero de salida, donde se guardarán todos los resultados generados por el programa. Enter output filename: A:\ACP\TIEMPOS.OUT
8. Introduce el nº de ejes necesarios para las representaciones gráficas:
9. Por último sal del programa con la opción QUIT.

Resultado del análisis proporcionado por MVSP

 ***** M V S P *****

MATRIZ DE CORRELACIONES CENTRADA

	Prof	Tran	Mena	Enfa	Cour	Toil	Repa	Somm	Tele	Lois
Prof	1.000	0.933	-0.908	-0.870	-0.658	-0.112	-0.455	-0.538	-0.059	-0.190
Tran	0.933	1	-0.869	-0.809	-0.503	-0.079	-0.613	-0.702	-0.044	-0.105
Mena	-0.908	-0.869	1.000	0.861	0.501	-0.035	0.361	0.433	-0.206	-0.113
Enfa	-0.870	-0.809	0.861	1.000	0.543	0.124	0.367	0.277	0.122	-0.109
Cour	-0.658	-0.503	0.501	0.543	1.000	0.593	-0.184	-0.030	0.216	0.235
Toil	-0.112	-0.079	-0.035	0.124	0.593	1.000	-0.360	-0.217	0.322	0.073
Repa	-0.455	-0.613	0.361	0.367	-0.184	-0.360	1.000	0.817	0.316	-0.040
Somm	-0.538	-0.702	0.433	0.277	-0.030	-0.217	0.817	1.000	0.018	0.208
Tele	-0.059	-0.044	-0.206	0.122	0.216	0.322	0.316	0.018	1	-0.095
Lois	-0.190	-0.105	-0.113	-0.109	0.235	0.073	-0.040	0.208	-0.095	1

EJES	AUTOVALORES	PORCENTAJE TOTAL	PORCENTAJE ACUMULADO
1	4.589	45.89	45.89
2	2.120	21.20	67.09
3	1.321	13.21	80.29
4	1.195	11.95	92.25
5	0.468	4.68	96.93
6	0.199	1.99	98.92
7	0.047	0.47	99.39
8	0.037	0.37	99.76
9	0.024	0.24	100.00
10	6.161E-09	6.161E-08	100.00

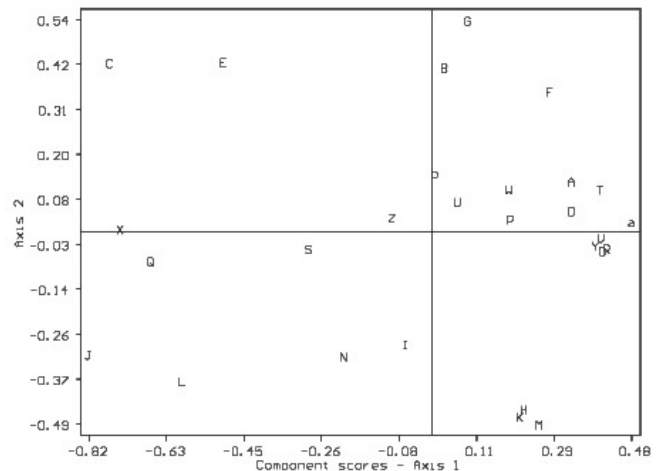
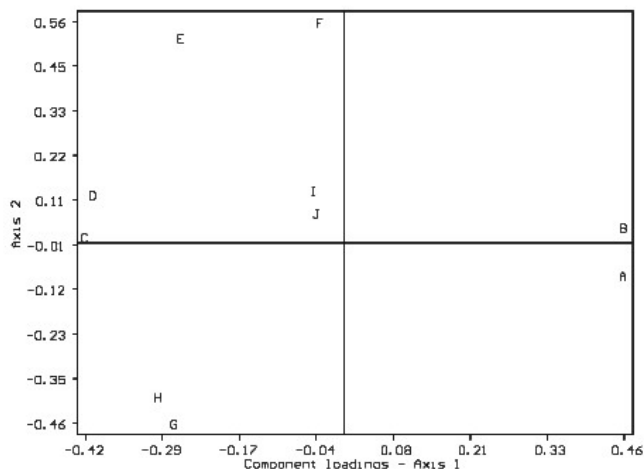
AUTOVECTORES (COMPONENTES LOADINGS)

	PLOT	AXIS1	AXIS2	AXIS3	AXIS4	AXIS5	AXIS6	AXIS7	AXIS8	AXIS9	AXIS10
Prof	A	0.456	-0.083	0.074	-0.061	0.140	-0.041	-0.067	0.489	0.095	0.708
Tran	B	0.457	0.040	0.007	-0.042	-0.162	0.019	0.304	-0.420	-0.685	0.150
Mena	C	-0.420	0.016	-0.315	-0.196	0.006	0.098	-0.065	-0.513	0.152	0.621
Enfa	D	-0.407	0.123	-0.073	-0.269	-0.278	-0.572	0.365	0.369	-0.243	0.095
Cour	E	-0.263	0.522	0.004	0.111	-0.125	0.610	-0.122	0.337	-0.345	0.101
Toil	F	-0.037	0.562	0.263	0.058	0.655	-0.362	-0.106	-0.169	-0.094	0.036
Repa	G	-0.275	-0.460	0.371	-0.013	0.004	-0.114	-0.590	0.002	-0.456	0.080
Somm	H	-0.301	-0.391	0.166	0.286	0.457	0.244	0.593	0.081	-0.106	0.093
Tele	I	-0.046	0.133	0.809	-0.138	-0.352	0.098	0.193	-0.174	0.300	0.123
Lois	J	-0.043	0.076	-0.026	0.876	-0.314	-0.275	-0.042	-0.072	0.059	0.200

VALORES DE LAS COMPONENTES PRINCIPALES

		PLOT	AXIS1	AXIS2	AXIS3	AXIS4	AXIS5	AXIS6	AXIS7	AXIS8	AXIS9	AXIS10
HAUS	A	0.335	0.130	0.354	-0.109	-0.161	0.126	-0.002	-0.020	-0.020	4.0E-06	
FAUS	B	0.032	0.419	0.125	-0.083	0.237	0.009	-0.011	-0.009	-0.025	-7.8E-07	
FNAU	C	-0.766	0.430	0.200	0.098	-0.196	-0.170	0.003	0.042	0.030	-1.0E-05	
HMUS	D	0.336	0.055	0.356	-0.139	-0.196	0.157	0.035	-0.016	9.4E-04	5.4E-06	
FMUS	E	-0.494	0.432	0.151	-0.020	-0.070	-0.108	-0.012	0.045	0.013	-1.4E-05	
HCUS	F	0.284	0.358	0.258	0.148	-0.067	0.151	-0.059	0.031	0.010	-8.3E-06	
FCUS	G	0.088	0.538	0.245	0.028	0.306	-0.024	-0.020	-0.062	0.001	2.4E-05	
HAWE	H	0.222	-0.447	0.211	0.009	0.044	-0.027	-0.028	0.017	0.009	-4.0E-06	
FAWE	I	-0.059	-0.283	0.051	-0.178	0.236	-0.002	-0.003	0.017	-0.018	-9.5E-06	
FNAW	J	-0.817	-0.309	0.168	0.027	-0.043	-0.085	0.006	-0.040	-0.065	2.7E-06	
HMWE	K	0.213	-0.466	0.243	-0.028	0.041	-0.047	-0.023	0.015	0.025	-4.6E-07	
FMWE	L	-0.592	-0.376	0.111	-0.139	-0.065	-0.002	-0.048	-0.019	-0.027	3.6E-06	
HCWE	M	0.259	-0.486	0.099	0.193	-0.054	0.025	0.016	0.034	0.023	-9.4E-06	
FCWE	N	-0.208	-0.313	0.103	0.283	0.270	0.021	-0.004	0.038	0.043	-6.7E-06	
HAYO	O	0.409	-0.046	-0.134	0.045	-0.061	-0.067	-0.024	9.5E-04	-0.048	-8.7E-06	
FAYO	P	0.190	0.034	-0.305	-0.403	-0.008	-0.045	-0.018	0.005	0.030	4.9E-06	
FNAY	Q	-0.669	-0.071	-0.309	0.100	-0.052	0.151	-0.036	-0.078	0.033	3.7E-05	
HMYO	R	0.420	-0.040	-0.091	0.021	-0.075	-0.057	-0.007	-0.010	-0.049	-5.1E-06	
FMYO	S	-0.291	-0.041	-0.306	-0.220	-0.083	0.035	-0.041	-0.033	0.036	2.1E-05	
HCYO	T	0.404	0.110	-0.304	0.412	-0.105	0.005	-0.048	0.050	-0.007	-1.9E-05	
FCYO	U	0.063	0.079	-0.282	0.194	0.023	-0.035	-0.060	-0.027	-0.001	1.1E-05	
HAES	V	0.406	-0.013	-0.025	-0.061	-0.028	-0.100	0.060	-0.030	-0.004	8.8E-06	
FAES	W	0.187	0.111	-0.258	-0.386	0.051	0.016	0.004	0.044	-0.001	-1.6E-05	
FNAE	X	-0.741	0.009	-0.127	0.184	0.000	0.122	0.112	0.022	-0.010	-1.3E-05	
HMES	Y	0.393	-0.032	0.080	-0.150	-0.041	-0.099	0.069	-0.053	0.066	3.1E-05	
FMES	Z	-0.093	0.039	-0.224	-0.380	0.008	0.055	0.031	0.065	-0.024	-2.9E-05	
HCES	a	0.478	0.028	-0.201	0.371	-0.067	-0.072	0.049	-0.037	-0.008	1.1E-05	
FCES	b	0.010	0.152	-0.189	0.183	0.159	0.070	0.058	0.009	-0.013	-7.6E-06	

figura 2: Variables e individuos



Datos transformados

$$x_{ij} = \frac{r_{ij} - \bar{r}_j}{\sigma_j}$$



	Prof	Tran	Mena	Enfa	Cour	Toil	Repa	Somm	Tele	Lois
1	0,716	1,142	-1,112	-0,779	0,350	0,007	-0,124	-0,887	1,953	-0,489
2	0,109	0,082	-0,138	-0,111	0,976	2,217	-0,719	-0,373	0,402	-0,648
3	-1,977	-1,825	1,117	2,564	1,914	1,333	0,470	-0,031	1,565	1,341
4	0,738	1,142	-1,087	-0,779	0,194	-0,435	-0,124	-0,715	2,082	-0,648
5	-1,219	-1,210	0,737	1,795	1,633	1,510	0,034	-0,339	1,126	0,433
6	0,603	0,612	-1,163	-1,114	1,289	0,891	-0,719	-0,887	1,307	0,625
7	0,141	0,167	-0,415	-0,512	1,007	3,101	-0,878	-0,373	0,842	-0,155
8	0,909	0,294	-0,933	-0,880	-1,619	-0,877	1,264	0,755	0,402	-0,250
9	0,271	-0,341	0,153	-0,111	-0,900	0,007	0,946	1,029	-0,321	-1,333
10	-1,933	-1,676	1,491	1,795	0,100	-0,435	2,454	1,953	0,661	0,354
11	0,922	0,231	-0,923	0,779	-1,776	-0,877	1,343	0,755	0,583	-0,394
12	-1,268	-1,358	1,286	1,193	-0,212	-1,053	2,216	1,303	0,506	-0,553
13	0,864	0,400	-1,051	-1,114	-1,463	-1,584	0,867	0,926	0,014	0,672
14	-0,096	-1,104	-0,077	-0,646	-0,524	0,183	1,145	2,158	-0,398	0,736
15	0,895	1,142	-0,805	-0,612	-0,743	-0,435	-0,521	-0,887	-0,760	0,306
16	0,491	0,400	0,501	0,390	-0,587	-0,435	-0,918	-1,400	-1,018	-1,763
17	-1,977	-1,613	2,219	0,725	1,132	-0,877	0,470	0,995	-1,018	0,545
18	0,895	1,248	-0,846	-0,612	-0,743	-0,435	-0,521	-0,887	-0,501	0,195
19	-0,855	-0,722	1,532	0,858	0,225	-0,877	-0,045	-0,373	-0,889	-0,808
20	0,738	0,824	-0,933	-1,114	0,194	-0,435	-1,315	-0,887	-1,535	2,058
21	-0,078	0,061	0,209	-0,345	0,100	0,095	-0,640	-0,407	-1,406	0,991
22	0,895	1,185	-0,794	-0,378	-1,025	-0,081	-0,719	-0,750	-0,088	-0,187
23	0,572	0,421	0,312	0,290	-0,086	-0,081	-1,037	-1,160	-0,915	-1,875
24	-1,915	-1,655	1,624	1,293	1,539	-0,258	0,391	1,850	-0,34	0,832
25	0,904	0,994	-0,733	-0,378	-1,275	-0,081	-0,640	-0,784	0,583	-0,569
26	-0,065	-0,150	0,799	0,892	0,319	-0,435	-0,442	-0,476	-0,682	-1,827
27	0,792	1,312	-1,071	-1,114	-0,649	-0,258	-1,275	-0,544	-1,070	1,867
28	-0,074	-0,002	0,102	-0,412	0,632	0,625	-0,957	0,447	-1,070	0,545

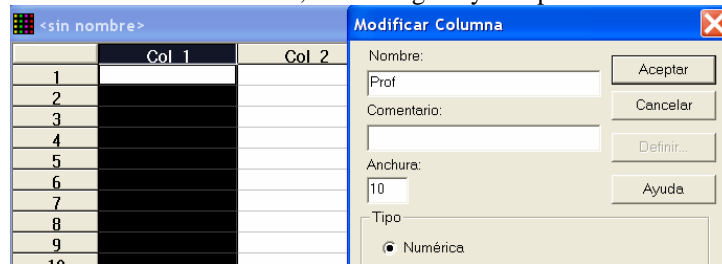
$$x_{ij} = \frac{r_{ij} - \bar{r}_j}{\sigma_j \sqrt{n}}$$


Prof	Tran	Mena	Enfa	Cour	Toil	Repa	Somm	Tele	Lois
0,135	0,215	-0,210	-0,147	0,066	0,001	-0,023	-0,1676	0,369	-0,092
0,020	0,015	-0,026	-0,020	0,184	0,419	-0,136	-0,0706	0,076	-0,122
-0,373	-0,344	0,211	0,484	0,361	0,251	0,088	-0,0059	0,295	0,253
0,139	0,215	-0,205	-0,147	0,036	-0,082	-0,023	-0,1353	0,393	-0,122
-0,230	-0,228	0,139	0,339	0,308	0,285	0,006	-0,0641	0,212	0,082
0,114	0,115	-0,219	-0,210	0,243	0,168	-0,136	-0,1676	0,247	0,118
0,026	0,031	-0,078	-0,096	0,190	0,586	-0,166	-0,0706	0,159	-0,029
0,171	0,055	-0,176	-0,166	-0,306	-0,165	0,238	0,1428	0,076	-0,047
0,051	-0,064	0,028	-0,020	-0,170	0,001	0,178	0,1945	-0,060	-0,252
-0,365	-0,316	0,281	0,339	0,019	-0,082	0,463	0,3691	0,124	0,066
0,174	0,043	-0,174	-0,147	-0,335	-0,165	0,253	0,1428	0,110	-0,074
-0,239	-0,256	0,243	0,225	-0,040	-0,199	0,418	0,2462	0,095	-0,104
0,163	0,075	-0,198	-0,210	-0,276	-0,299	0,163	0,1751	0,002	0,127
-0,018	-0,208	-0,014	-0,122	-0,099	0,034	0,216	0,4079	-0,075	0,139
0,169	0,215	-0,152	-0,115	-0,140	-0,082	-0,098	-0,1676	-0,143	0,057
0,092	0,075	0,094	0,073	-0,111	-0,082	-0,173	-0,2646	-0,192	-0,333
-0,373	-0,304	0,419	0,137	0,214	-0,165	0,088	0,1880	-0,192	0,103
0,169	0,235	-0,159	-0,115	-0,140	-0,082	-0,098	-0,1676	-0,094	0,036
-0,161	-0,136	0,289	0,162	0,042	-0,165	-0,008	-0,0706	-0,168	-0,152
0,139	0,155	-0,176	-0,210	0,036	-0,082	-0,248	-0,1676	-0,290	0,388
-0,014	0,011	0,039	-0,065	0,019	0,018	-0,121	-0,0770	-0,265	0,187
0,169	0,223	-0,150	-0,071	-0,193	-0,015	-0,136	-0,1417	-0,016	-0,035
0,108	0,079	0,059	0,054	-0,016	-0,015	-0,196	-0,2193	-0,173	-0,354
-0,361	-0,312	0,307	0,244	0,290	-0,048	0,073	0,3497	-0,065	0,157
0,170	0,187	-0,138	-0,071	-0,241	-0,015	-0,121	-0,1482	0,110	-0,107
-0,012	-0,028	0,151	0,168	0,060	-0,082	-0,083	-0,0900	-0,129	-0,345
0,149	0,247	-0,202	-0,210	-0,122	-0,048	-0,241	-0,1029	-0,202	0,352
-0,014	-0,001	0,019	-0,077	0,119	0,118	-0,181	0,0845	-0,202	0,103

Análisis de componentes principales con Statgraphics 5.1. Práctica 1.



Pasos a realizar:

1. Abrir el programa Statgraphic 5.1
2. Elegir la opción  Analizar Datos Existentes o Introducir Nuevos Datos...
3. Elegir la opción  Deseo Introducir Nuevos Datos...
4. Introducir los nombres de las 10 variables, con 10 dígitos y de tipo numérico:



5. Ir a la dirección <http://www.ugr.es/~gallardo/pdf/prac-acp.pdf>
6. Marcar la tabla de datos
7. Una vez seleccionada elegir con el botón derecho del ratón la opción  Abrir tabla en hoja de cálculo
8. Marcar en la hoja de cálculo los datos numéricos, copiarlos y pegarlos en los datos del Statgraphics
9. Elegir el menú **Avanzado** y **Métodos Multivariantes** ▶ **Componentes Principales...**
10. Elegir las 10 variables, sin ordenar



11. Marcar en  las 3 opciones tabulares
12. Marcar en  las 7 opciones gráficas
13. Si no salen los gráficos de dos o tres dimensiones, pulsar con el botón derecho del ratón en cualquier gráfico o resultado numérico, y elegir **Opciones de Análisis...** **Números de Componentes:**
14. Elegir la opción: Número de componentes y marcar 4:
15. El resultado final debe parecerse a esta imagen:

Resumen del Análisis

Datos/Variabes:

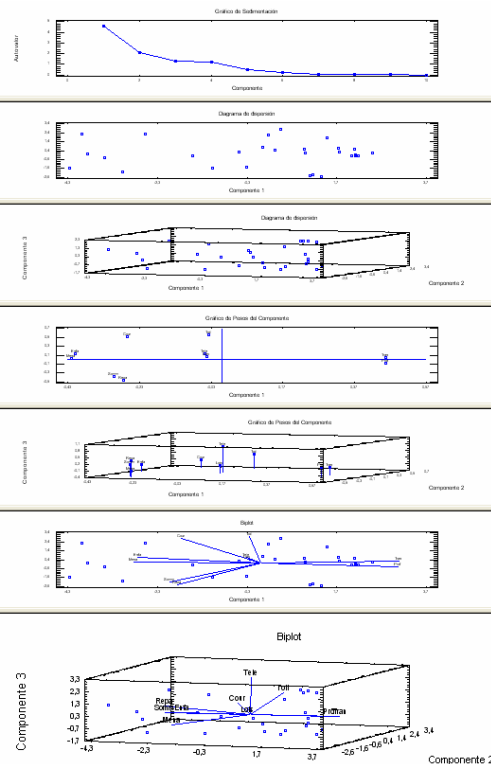
Prof
Tran
Mena
Enfa
Cour
Toil
Repa
Somm

Tabla de Pesos de los Componentes

	Componentes 1	Componentes 2	Componentes 3	Componentes 4
Prof	0,456172	-0,0831378	0,073587	-0,0612
Tran	0,457388	0,0399155	0,00730326	-0,0416
Mena	-0,4201	0,0155579	-0,315342	-0,1953
Enfa	-0,40712	0,122649	-0,0728517	-0,2693
Cour	-0,2631	0,522412	0,00396746	0,11071
Toil	-0,0371177	0,56189	0,262902	0,05812

Tabla de Componentes Principales

Fila	Componentes 1	Componentes 2	Componentes 3	Componentes 4
1	1,74099	0,673691	1,83755	-0,564791
2	0,168497	2,1754	0,648857	-0,429672
3	-3,98036	2,23666	1,04141	0,510938
4	1,74731	0,287398	1,85117	-0,71976
5	-2,56715	2,24412	0,782865	-0,105665
6	1,47571	1,85764	1,33848	0,768171
7	0,456859	2,79305	1,27305	0,144987
8	1,15515	-2,32502	1,09651	0,0449763
9	-0,306382	-1,46833	0,267483	-0,92631



Resultado del análisis proporcionado por Statgraphics

Resumen del Análisis

Número de componentes extraídos: 4

Datos/Variabes:

Análisis de Componentes Principales

	Componente		Porcentaje de	Acumulado
	Número	Autovalor	Varianza	Porcentaje
Prof	1	4,58867	45,887	45,887
Tran	2	2,11984	21,198	67,085
Mena	3	1,32098	13,210	80,295
Enfa	4	1,19526	11,953	92,247
Cour	5	0,468411	4,684	96,932
Toil	6	0,199047	1,990	98,922
Repa	7	0,0468132	0,468	99,390
Somm	8	0,0370651	0,371	99,761
Tele	9	0,0239189	0,239	100,000
Lois	10	2,67083E-16	0,000	100,000

Entrada de datos: observaciones

Número de casos completos: 28

Tratamiento de valor perdido: lista

considerada, Estandarizado: si

El StatAdvisor

Este procedimiento realiza un análisis de componentes principales. El propósito del análisis es obtener un pequeño número de combinaciones lineales de las 10 variables que explican la mayoría de la variabilidad en los datos. En este caso, se han extraído 4 componentes como se ha pedido. Juntos explican el 92,2474% de la variabilidad en los datos originales.

El StatAdvisor

Este procedimiento realiza un análisis de componentes principales. El propósito del análisis es obtener un pequeño número de combinaciones lineales de las 10 variables que explican la mayoría de la variabilidad en los datos. En este caso, se han extraído 4 componentes como se ha pedido. Juntos explican el 92,2474% de la variabilidad en los datos originales.

Tabla de Pesos de los Componentes

	Componentes 1	Componentes 2	Componentes 3	Componentes 4
Prof	0,456172	-0,0831378	0,073587	-0,0612328
Tran	0,457388	0,0399155	0,00730326	-0,0416677
Mena	-0,4201	0,0155579	-0,315342	-0,195895
Enfa	-0,40712	0,122649	-0,0728517	-0,269325
Cour	-0,2631	0,522412	0,00396746	0,110711
Toil	-0,0371177	0,56189	0,262902	0,0581298
Repa	-0,274653	-0,459739	0,370916	-0,0129315
Somm	-0,30072	-0,391013	0,166054	0,285831
Tele	-0,0464201	0,132622	0,809201	-0,138342
Lois	-0,0430303	0,0757267	-0,0262929	0,875764

El StatAdvisor

Esta tabla muestra las ecuaciones de los componentes principales.
 Por ejemplo, el primer componente principal tiene la ecuación

$$0,456172*Prof + 0,457388*Tran - 0,4201*Mena - 0,40712*Enfa - \\ 0,2631*Cour - 0,0371177*Toil - 0,274653*Repa - 0,30072*Somm - \\ 0,0464201*Tele - 0,0430303*Lois$$

donde los valores de las variables en la ecuación están estandarizados
 substrayendo sus medias y dividiéndolos por sus desviaciones típicas.

Tabla de Componentes Principales

Fila	Componentes 1	Componentes 2	Componentes 3	Componentes 4
1	1,74099	0,673691	1,83755	-0,564791
2	0,168497	2,1754	0,648857	-0,429672
3	-3,98036	2,23666	1,04141	0,510938
4	1,74731	0,287398	1,85117	-0,71976
5	-2,56715	2,24412	0,782865	-0,105665
6	1,47571	1,85764	1,33848	0,768171
7	0,456859	2,79305	1,27305	0,144987
8	1,15515	-2,32502	1,09651	0,0449763
9	-0,306382	-1,46833	0,267483	-0,92631
10	-4,24548	-1,60315	0,874222	0,141205
11	1,1051	-2,41952	1,2624	-0,147543
12	-3,07485	-1,95309	0,577586	-0,721319
13	1,34534	-2,52562	0,516831	1,0044
14	-1,07931	-1,62532	0,533548	1,46871
15	2,12373	-0,236703	-0,696125	0,23502
16	0,986675	0,176823	-1,58662	-2,09392
17	-3,47369	-0,370284	-1,60586	0,519999
18	2,18118	-0,20781	-0,474455	0,107623
19	-1,51222	-0,21218	-1,5922	-1,14476
20	2,0968	0,570157	-1,58077	2,1383
21	0,32976	0,410697	-1,46369	1,00644
22	2,10812	-0,0681732	-0,128855	-0,315791
23	0,969918	0,574883	-1,34076	-2,00319
24	-3,84808	0,048557	-0,659749	0,957829
25	2,03995	-0,16741	0,41746	-0,778052
26	-0,481588	0,200298	-1,16258	-1,97625
27	2,48387	0,144169	-1,04339	1,92806
28	0,0541554	0,789053	-0,984357	0,950373

El StatAdvisor

Esta tabla muestra los valores de los componentes principales por cada fila del fichero de datos. Seleccione Peso de los Componentes de la lista de Opciones Tabulares, para obtener las ecuaciones para cada componente. Seleccione Diagrama de dispersión 2D o Diagrama de dispersión 3D en la lista de Opciones Gráficas para trazar este dato. Puede guardar los componentes pulsando el cuarto botón de la izquierda en la barra de herramientas del análisis.

Diagrama de dispersión

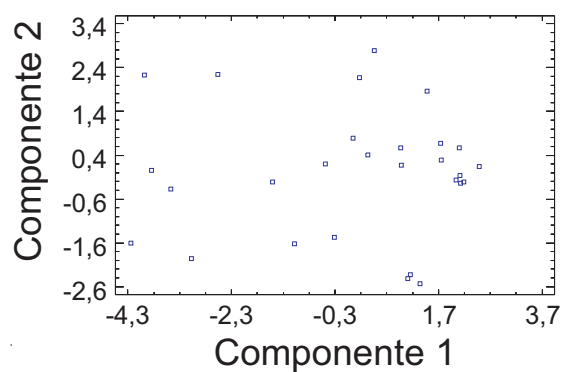
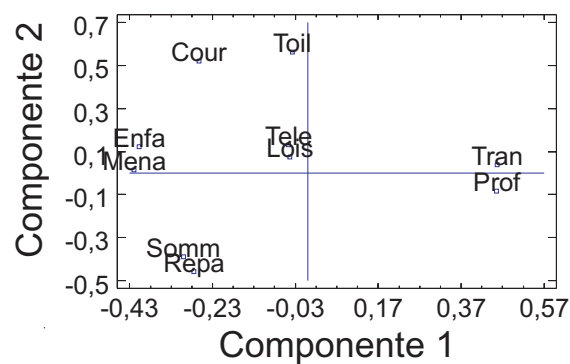


Gráfico de Pesos del Componente



Biplot

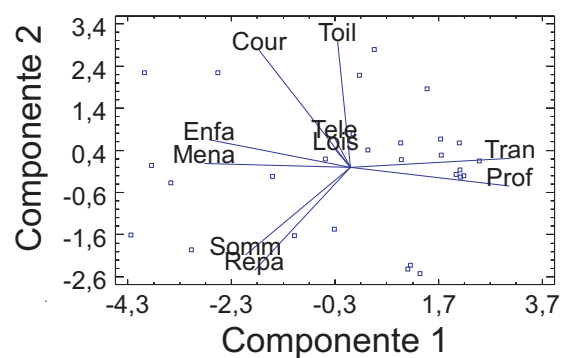


Diagrama de dispersión

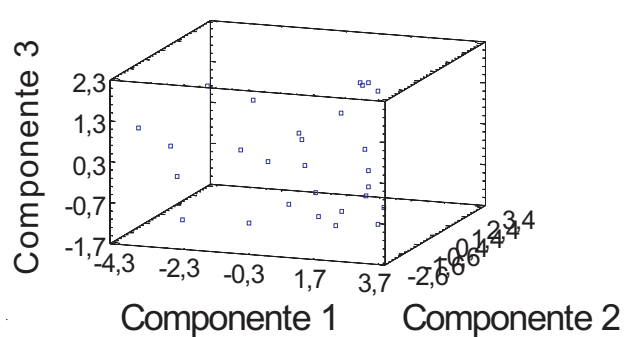
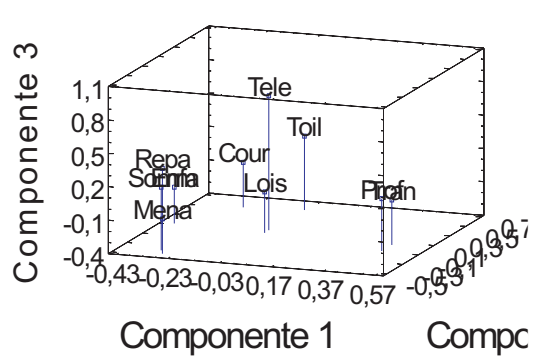
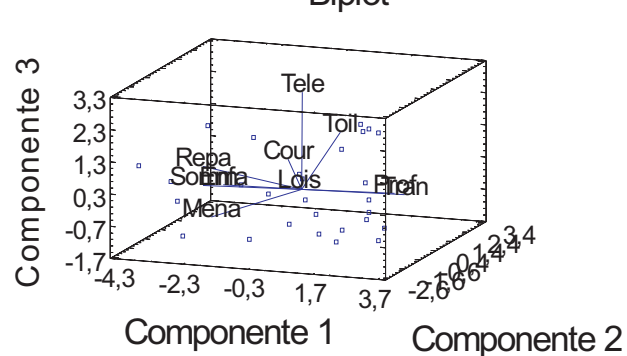


Gráfico de Pesos del Compore



Biplot



2ª Práctica de ACP

La siguiente tabla nos muestra el gasto medio de algunos tipos de familias clasificadas por el número de hijos: 2, 3, 4, y 5, y por una característica profesional del cabeza de familia: Trabajador manual (TM), Empleado no manual (EN) y Directivo (DI), el número de hijos aparece en la tercera columna de la denominación del cabeza de familia. Realiza un Análisis de Componentes Principales, utilizando los programas Bmdp, Mvsp y Statgraphic.

	Pan	Legumbres	Frutas	Carnes	Aves	Leche	Vino
TM2	332	428	354	1437	526	247	427
EN2	293	559	388	1527	567	239	258
DI2	372	767	652	1948	927	235	433
TM3	406	563	341	1507	544	324	407
EN3	386	608	396	1501	558	319	363
DI3	438	843	689	2345	1148	243	341
TM4	534	660	367	1620	638	414	407
EN4	460	699	484	1856	762	400	416
DI4	385	789	621	2366	1149	304	282
TM5	655	776	423	1848	759	495	486
EN5	584	995	548	2056	893	518	319
DI5	515	1097	887	2630	1167	651	284

Resultado del análisis proporcionado por Statgraphics

- En el resumen tabular se obtienen los autovalores correspondientes:

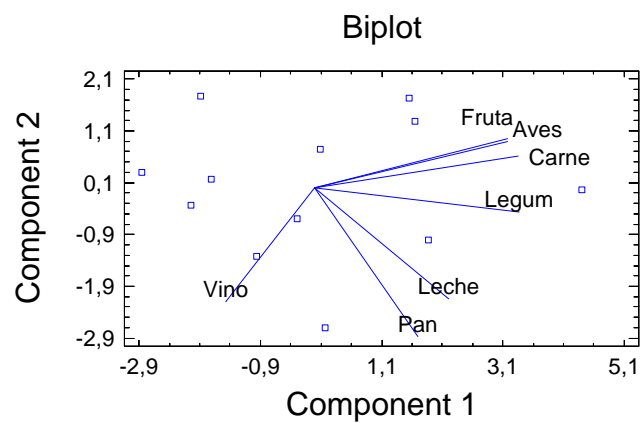
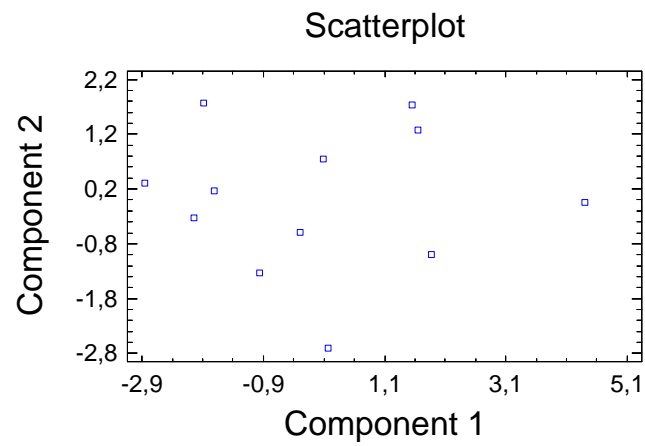
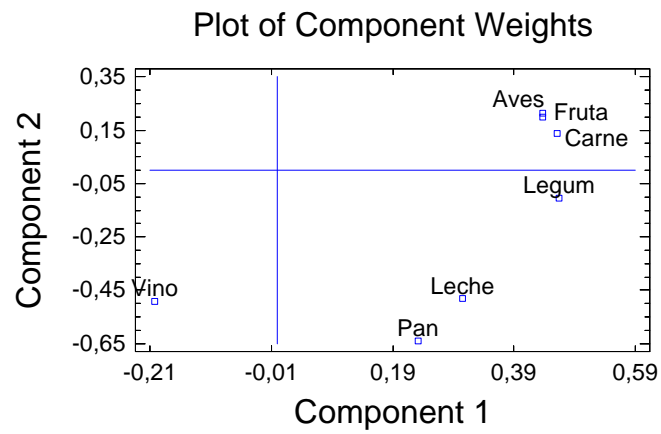
Valores propios	% de Varianza	% Acumulado
4.35627	62.232	62.232
1.7252	24.646	86.878
0.659858	9.427	96.305

- Las coordenadas de las variables respecto a los ejes:

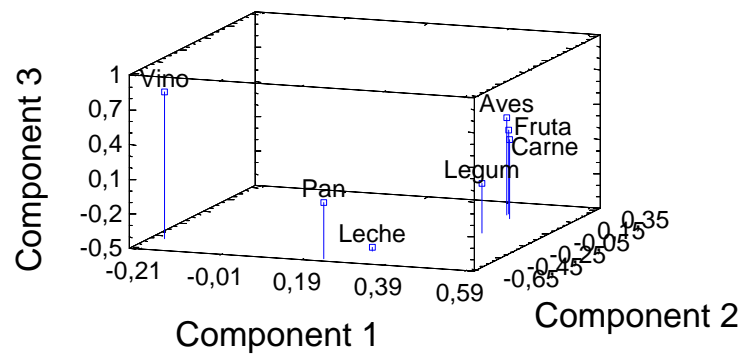
	Componente 1ª	Componente 2ª	Componente 3ª
Pan	0.232726	-0.61118	-0.0125995
Legum	0.464917	-0.104514	-0.065775
Fruta	0.437054	0.214587	0.222318
Carne	0.461599	0.137908	0.183188
Aves	0.437879	0.198324	0.337175
Leche	0.304664	-0.480278	-0.46252
Vino	-0.20259	-0.492849	0.764801

- Por último se obtienen las coordenadas de los individuos respecto a los ejes:

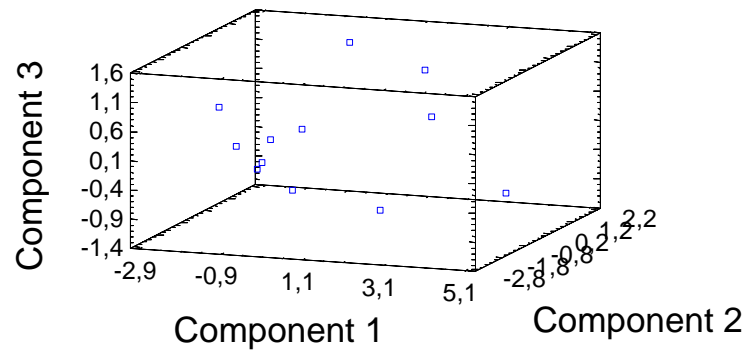
Filas	Componente 1ª	Componente 2ª	Componente 3ª
TM2	-2.85152	0.304398	0.364216
EN2	-1.89125	1.76151	-1.30801
DI2	0.0903039	0.753002	1.51555
TM3	-2.04651	-0.331288	-0.132408
EN3	-1.70757	0.162182	-0.509014
DI3	1.6444	1027957	1.00412
TM4	-0.960462	-1.32871	-0.28064
EN4	-0.288971	-0.590146	0.288928
DI4	1.55405	1.73373	0.110602
TM5	0.172198	-2.70525	0.5673
EN5	1.87946	-1.00091	-0.918834
DI5	4.40587	-0.0380711	-0.701817



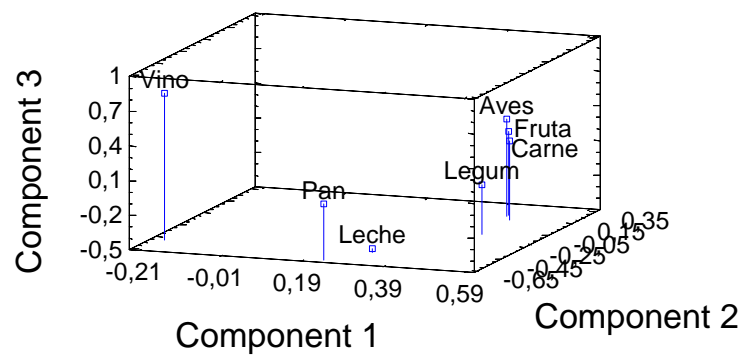
Plot of Component Weights

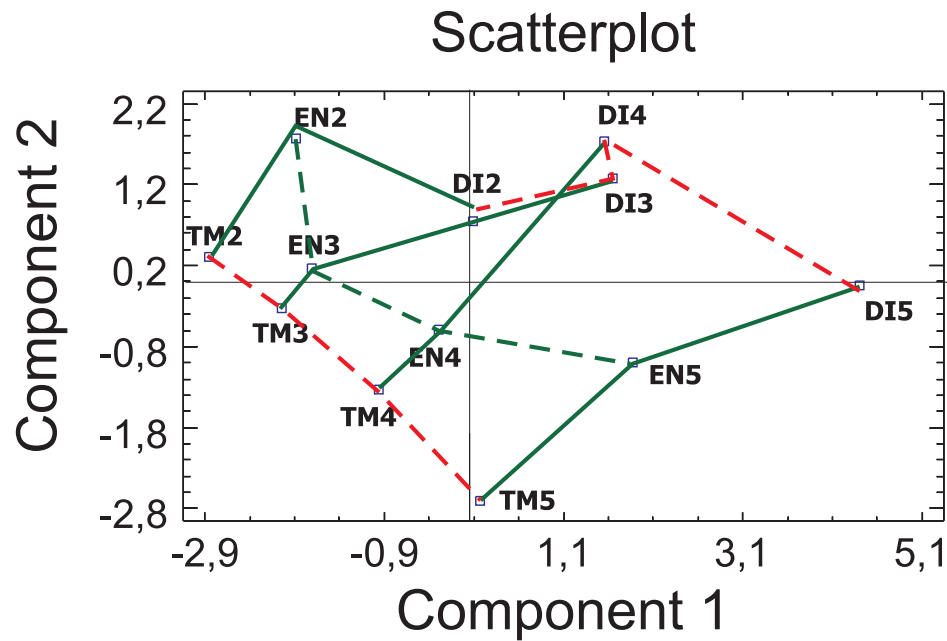


Scatterplot



Plot of Component Weights





Interpretación

- El análisis de la nube de los individuos se hace respecto a su centro de gravedad, no ocurre lo mismo para las variables. La simple consulta de sus coordenadas sobre el primer eje, nos muestra que con excepción del vino, todos los productos están en el mismo lado respecto al origen. Tal situación traduce el hecho de que la mayoría de las variables están correlacionadas positivamente entre ellas, si para un individuo una variable toma un valor fuerte, todas las demás variables toman un valor fuerte. Esta característica se llama factor de tamaño, y aquí el factor de tamaño opone a los grandes consumidores respecto a los pequeños consumidores.
- En el gráfico de los individuos, si unimos los puntos correspondientes a un mismo número de hijos por un lado, y a la misma característica profesional por otro, observamos que el factor de tamaño, puesto en evidencia en la representación de las variables, está relacionado con el tamaño global de la familia, y por otro lado con su nivel socioprofesional: el gasto global aumenta con el número de hijos y nivel social.
- La superposición de las representaciones gráficas de los individuos y las variables, con las precauciones de interpretación adecuadas, (ver ambas representaciones biplot), da una visión más completa, sugiriendo qué variables son responsables de la proximidad entre puntos. Se observa por ejemplo, que el total de gastos en pan, leche y sobre todo en vino, se sitúa cerca de los empleados manuales y no manuales y lejos de los directivos.

Ejercicio propuesto de ACP

La siguiente tabla, recogida del Anuario El País (año 1999, pag 450), nos muestra para las 8 provincias andaluzas los datos sobre población, extensión y varios indicadores económicos. Realiza un Análisis de Componentes Principales utilizando los programas Statgraphics y Mvsp, presentando los resultados obtenidos, las representaciones gráficas, las diferencias obtenidas por ambos programas y las conclusiones obtenidas.

	Población	Extensión (km ²)	Cuota de mercado	Indice turístico	Teléf.	Automóv.	Oficinas bancarias	Actividad económica
Almería	501761	8775	1266	1300	158630	171431	456	918
Cádiz	1105762	7440	2402	1999	315734	342988	609	1599
Córdoba	761401	13772	1787	716	233488	220020	598	1263
Granada	808053	12633	1908	1712	265672	267780	643	1242
Huelva	454735	10128	1055	756	132071	126388	365	776
Jaén	648551	13490	1568	369	188132	166351	541	1014
Málaga	1249290	7306	2947	5638	435823	462001	843	2475
Sevilla	1705320	14036	3839	2881	540018	538280	1183	3119