

ANALISIS DE COMPONENTES PRINCIPALES

Dr. Porfirio Gutiérrez González

ANÁLISIS DE COMPONENTES PRINCIPALES

- **El análisis de componentes principales (ACP) es una de las técnicas estadísticas más utilizadas para el análisis de datos multivariados, ya que nos permite verificar las propiedades de la matriz de datos desde distintas perspectivas.**
- **La idea básica del ACP es expresar las variables originales en un número reducido de nuevas variables $k < p$, permitiendo con esto disminuir la dimensionalidad del problema.**
- **Generar nuevas variables con propiedades convenientes que expresen la información contenida en un conjunto de datos.**
- **Las nuevas variables son independientes y de varianza máxima.**

DEFINICIÓN DE COMPONENTES PRINCIPALES

- ❑ El primer componente es la combinación lineal de las variables originales, con varianza máxima.
- ❑ El segundo componente es la combinación lineal de las variables originales, independiente del primer componente, con la varianza máxima.
- ❑ El tercer componente es la combinación lineal de las variables originales, con la maximiza varianza independiente de los dos primeros componente.
- ❑ Así sucesivamente, el K componente es la combinación lineal de las variables originales, con varianza máxima e independiente de los $k-1$ componentes anteriores.

Sean X_1, X_2, \dots, X_p las variables originales y sean Y_1, Y_2, \dots, Y_p los componentes principales. Esto expresado matemáticamente como

$$Y_1 = \alpha_{11}X_1 + \alpha_{12}X_2 + \alpha_{13}X_3 + \dots + \alpha_{1p}X_p$$

$$Y_2 = \alpha_{21}X_1 + \alpha_{22}X_2 + \alpha_{23}X_3 + \dots + \alpha_{2p}X_p$$

$$Y_3 = \alpha_{31}X_1 + \alpha_{32}X_2 + \alpha_{33}X_3 + \dots + \alpha_{3p}X_p$$

$$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$$

$$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$$

$$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$$

$$Y_p = \alpha_{p1}X_1 + \alpha_{p2}X_2 + \alpha_{p3}X_3 + \dots + \alpha_{pp}X_p$$

Donde X_1, X_2, \dots, X_p son las variables originales

$$\text{Var}(Y_1) = \lambda_1 \quad \text{Var}(X_1) = S_1$$

$$\text{Var}(Y_2) = \lambda_2 \quad \text{Var}(X_2) = S_2$$

$$\text{Var}(Y_3) = \lambda_3 \quad \text{Var}(X_3) = S_3$$

$$\text{Var}(Y_p) = \lambda_p \quad \text{Var}(X_p) = S_p$$

α_{ij} pesos o cargas

DERIVACIÓN ALGEBRAICA DE LOS COMPONENTES PRINCIPALES

Para derivar los componentes principales se tienen procesos algebraicos definidos. Sean X_1, X_2, \dots, X_p variables aleatorias que constituyen la matriz de datos \mathbf{X} y sea \mathbf{S} la matriz de varianzas y covarianzas.

El primer componente principal está dado por

$$Y_1 = \alpha_{11}X_1 + \alpha_{12}X_2 + \alpha_{13}X_3 + \dots + \alpha_{1p}X_p$$

$$Y_1 = \alpha'_1 X$$

$$\text{Var}\{Y_1\} = \text{Var}\{\alpha'_1 X\} = \alpha'_1 \mathbf{S} \alpha_1$$

Siendo esta la máxima varianza, donde

$$\alpha'_1 = [\alpha_{11}, \alpha_{12}, \alpha_{13}, \dots, \alpha_{1p}]$$

La única restricción del sistema es que $\alpha_1' \alpha_1 = 1$, ya que para maximizar $\alpha_1' \mathbf{S} \alpha_1$ basta con hacer crecer a α_1 y automáticamente la varianza Y_1 crecerá. Viéndolo de manera vectorial, lo que se busca no es alargar el vector, ya que esto se hace fácilmente con escalares. La idea es encontrar la dirección del vector \mathbf{Y}_1 y encontrar el α_1 que cumpla con la característica $\alpha_1' \alpha_1 = 1$.

La maximización de $\alpha_1' \mathbf{S} \alpha_1$ se resuelve con multiplicadores de Lagrange. Sea una función

$$\varphi_1 = \alpha_1' \mathbf{S} \alpha_1 - \lambda (\alpha_1' \alpha_1 - 1)$$

donde λ es un multiplicador de Lagrange. Derivando respecto a α_1 y después igualando a cero

$$\mathbf{S} \alpha_1 - \lambda \alpha_1 = 0$$

$$\mathbf{S} \alpha_1 = \lambda_1 \alpha_1$$

Implica que α_1 es un vector propio de la matriz \mathbf{S} y que λ_1 su correspondiente valor propio.

Para determinar que valor propio de \mathbf{S} es la solución multiplicando por la izquierda α_1'

$$\alpha_1' \mathbf{S} \alpha_1 = \alpha_1' \lambda \alpha_1$$

$$\alpha_1' \mathbf{S} \alpha_1 = \lambda \alpha_1' \alpha_1 = \lambda_1$$

Por lo que λ_1 es la varianza de Y_1 .

Como Y_1 es el primer componente que explica la mayor varianza entonces λ_1 es el mayor propio de la matriz \mathbf{S} . Su vector asociado α_1 contiene los pesos o cargas de cada variable original en el primer componente.

En este caso, ya que λ_1 es el máximo entonces el primer componente está dado por la combinación lineal

$$Y_1 = \alpha_{11}X_1 + \alpha_{12}X_2 + \alpha_{13}X_3 + \dots + \alpha_{1p}X_p$$

donde $\alpha_{11}, \alpha_{12}, \alpha_{13}, \dots, \alpha_{1p}$ son los coeficientes del vector propio correspondiente al mayor valor propio λ_1 de la matriz S , además la $\text{Var}(Y_1) = \alpha_1' S \alpha_1 = \lambda_1$.

Para el segundo componente principal estará dado por

$$Y_2 = \alpha_{21}X_1 + \alpha_{22}X_2 + \alpha_{23}X_3 + \dots + \alpha_{2p}X_p$$

El problema se centra en encontrar un vector $\alpha'_2 = [\alpha_{21}, \alpha_{22}, \alpha_{23}, \dots, \alpha_{2p}]$ tal que

$$\text{Var}(Y_2) = \alpha'_2 S \alpha_2 \quad \text{Sea máxima independiente de } Y_1$$

Con las restricciones $\alpha'_2 \alpha_2 = 1$ y $\alpha'_1 \alpha_2 = 0$ Ortogonalidad de Y_1 y Y_2

$$\text{Cov}(Y_1, Y_2) = \text{Cov}(\alpha'_1 X, \alpha'_2 X) = \alpha'_1 \text{Cov}(X' X) \alpha'_2 = \alpha'_1 S \alpha'_2$$

$$\text{Cov}(Y_1, Y_2) = \alpha_1 \alpha'_1 S \alpha_1 \alpha'_2 = \alpha_1 \lambda_1 \alpha'_2 = \lambda_1 \alpha_1 \alpha'_2 = 0 \quad \text{pero} \quad \lambda_1 \neq 0$$

$$\text{Por lo tanto} \quad \alpha_1 \alpha'_2 = 0$$

Por medio de los multiplicadores de Lagrange se maximiza de forma que

$$\varphi_2 = \alpha_2' \mathbf{S} \alpha_2 - \lambda (\alpha_2' \alpha_2 - 1) - \phi \alpha_2' \alpha_1$$

Derivando esto respecto a α_2 y luego igualando a cero y simplificando, se obtiene

$$\mathbf{S} \alpha_2 - \lambda \alpha_2 - \phi \alpha_1 = \mathbf{0}$$

Multiplicando a la izquierda por α_1' tenemos

$$\alpha_1' \mathbf{S} \alpha_2 - \alpha_1' \lambda \alpha_2 - \alpha_1' \phi \alpha_1 = 0$$

$$\alpha_1' \mathbf{S} \alpha_2 - \lambda \alpha_1' \alpha_2 - \phi \alpha_1' \alpha_1 = 0$$

$$\mathbf{Cov}(Y_1, Y_2) = \alpha_1' \mathbf{S} \alpha_2 = \lambda \alpha_1' \alpha_2 = 0$$

$$\alpha_1' \alpha_1 = 1, \quad \alpha_1' \alpha_2 = 0, \quad \phi = 0,$$

$$\mathbf{S} \alpha_2 - \lambda \alpha_2 = \mathbf{0} \quad \longrightarrow \quad \mathbf{S} \alpha_2 = \lambda \alpha_2 \quad \longrightarrow \quad \alpha_2' \mathbf{S} \alpha_2 = \alpha_2' \lambda \alpha_2 = \lambda_2 \alpha_2' \alpha_2 = \lambda_2$$

$$\mathbf{Var}(Y_2) = \lambda_2$$

Si este procedimiento se sigue entonces se conseguirá un tercer, un cuarto, un n-ésimo componente principal con los vectores propios $\alpha_3 \alpha_4 \alpha_5 \dots , \alpha_p$ correspondientes a los valores propios tercero, cuarto, etc. hasta el p-ésimo valor más grande de la matriz \mathbf{S} . La varianza de \mathbf{Y}_k es igual al k-ésimo valor propio λ_k . Esto es

$$\mathbf{Var} (\mathbf{Y}_k) = \mathbf{Var} (\alpha'_k \mathbf{X}) = \alpha'_k \mathbf{S} \alpha_k = \alpha'_k \lambda_k \alpha_k = \lambda_k$$

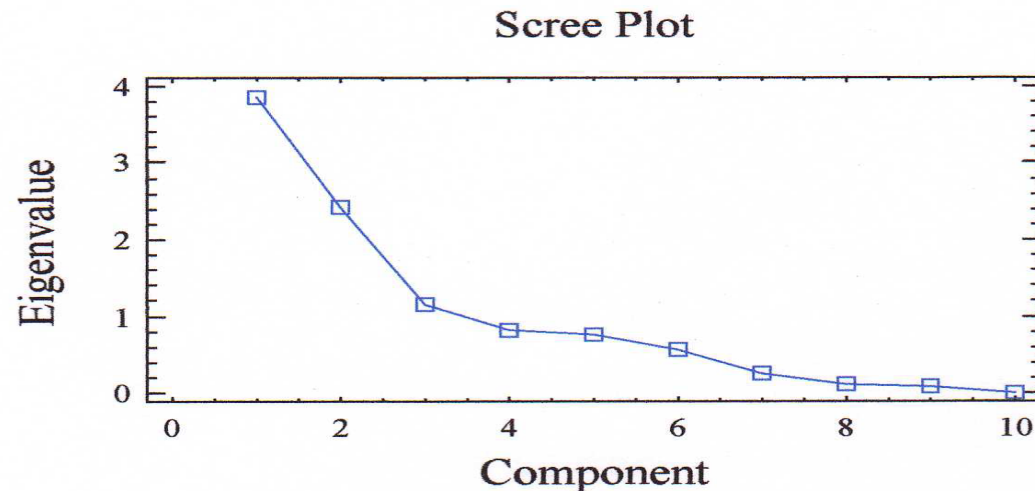
Existen dos métodos para ayudar a elegir el número de componentes principales:

Método 1: Supóngase que se desea tomar en cuenta el 100% de la variabilidad de los datos en las variables originales, entonces considerando

$$\textit{Varianza total} = [(\lambda_1 + \lambda_2 + \lambda_3 + \dots + \lambda_p) / (S_1 + S_2 + S_3 + \dots + S_p)] * 100$$

Se selecciona el número de componentes principales hasta cubrir una proporción adecuada de la variabilidad total de los datos.

Método 2: Se utiliza una grafica de sedimentación de los eigenvalores. En una grafica de sedimentación se utilizan parejas $(1, \lambda_1), (1, \lambda_2), \dots, (1, \lambda_p)$. Cuando los puntos de la grafica tienden a nivelarse, estos eigenvalores suelen estar suficientemente cercanos a cero, es probable que el correspondiente componente principal este midiendo muy poca información y no sea necesario interpretarse.

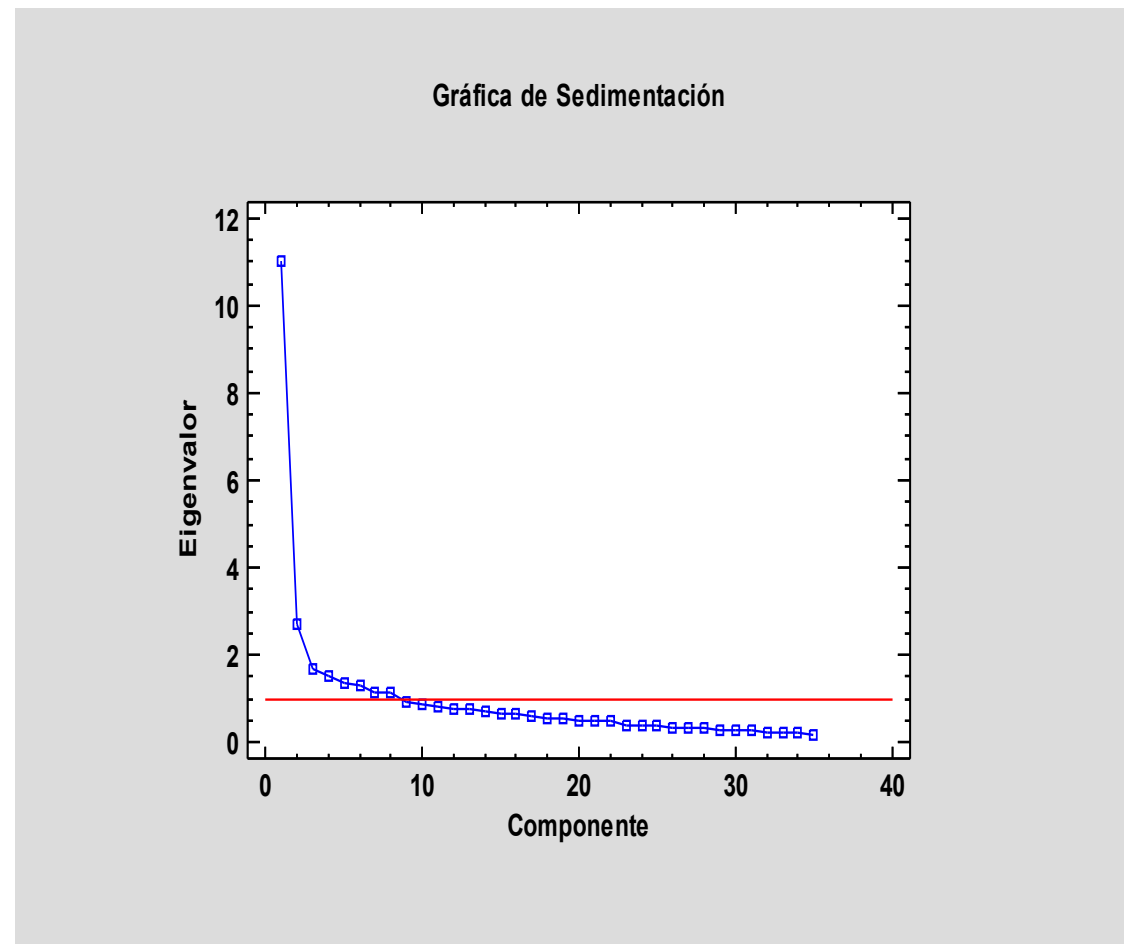


- Escala de Depresión del Centro de Estudios Epidemiológicos, Versión Revisada: CES-D-R
- 1. Tenía poco apetito
- 2. No podía quitarme la tristeza
- 3. Tenía dificultad para mantener mi mente en lo que estaba haciendo
- 4. Me sentía deprimido
- 5. Dormía sin descansar
- 6. Me sentía triste
- 7. No podía seguir adelante
- 8. Nada me hacía feliz
- 9. Sentía que era una mala persona
- 10. Había perdido interés en mis actividades diarias
- 11. Dormía más de lo habitual
- 12. Sentía que me movía muy lento
- 13. Me sentía agitado(a)
- 14. Sentía deseos de estar muerto
- 15. Quería hacerme daño
- 16. Me sentía cansado(a) todo el tiempo
- 17. Estaba a disgusto conmigo mismo(a)
- 18. Perdí peso sin intentarlo

- 19. Me costaba mucho trabajo dormir
- 20. Era difícil concentrarme en las cosas importantes
- 21. Me molestaba por cosas que usualmente no me molestan
- 22. Sentí que era tan bueno(a) como otra gente
- 23. Sentí que todo lo que hacía era con esfuerzo
- 24. Me sentía esperanzado(a) hacia el futuro
- 25. Pensé que mi vida ha sido un fracaso
- 26. Me sentía temeroso(a)
- 27. Me sentía feliz
- 28. Hablé menos de lo usual
- 29. Me sentía solo(a)
- 30. Las personas eran poco amigables
- 31. Disfruté de la vida
- 32. Tenía ataques de llanto
- 33. Me divertí mucho
- 34. Sentí que iba a darme por vencido(a)
- 35. Sentí que le desagradaba a la gente

Análisis de componentes principales datos estandarizados

Número de componente	Eigenvalor	Porcentaje de Varianza	Porcentaje Acumulado
1	10.9981	31.423	31.423
2	2.71605	7.76	39.183
3	1.6657	4.759	43.942
4	1.50752	4.307	48.25
5	1.35486	3.871	52.121
6	1.31994	3.771	55.892
7	1.16608	3.332	59.224
8	1.12077	3.202	62.426
9	0.943754	2.696	65.122
10	0.870517	2.487	67.609
11	0.828141	2.366	69.976
12	0.782146	2.235	72.21
13	0.739526	2.113	74.323
14	0.707103	2.02	76.343
15	0.655455	1.873	78.216
16	0.634485	1.813	80.029
17	0.592491	1.693	81.722
18	0.540291	1.544	83.266
19	0.526332	1.504	84.769
20	0.517287	1.478	86.247
21	0.477708	1.365	87.612
22	0.474663	1.356	88.968
23	0.409855	1.171	90.139
24	0.405945	1.16	91.299
25	0.398501	1.139	92.438
26	0.353251	1.009	93.447
27	0.333692	0.953	94.4
28	0.31795	0.908	95.309
29	0.296629	0.848	96.156
30	0.270001	0.771	96.928
31	0.256478	0.733	97.661
32	0.247347	0.707	98.367
33	0.216301	0.618	98.985
34	0.210256	0.601	99.586
35	0.14488	0.414	100



	Componente 1	Componente 2	Componente 3	Componente 4	Componente 5	Componente 6	Componente 7	Componente 8
PREG1	0.125494	0.0588908	-0.21012	0.0507131	0.295933	-0.0650813	-0.132107	-0.291842
PREG2	0.22011	-0.0370008	0.00139899	0.348647	0.0162863	-0.0304575	-0.0570047	-0.0250237
PREG3	0.211156	0.150142	-0.164516	-0.108663	0.0384367	-0.100285	-0.00580277	0.131541
PREG4	0.217712	-0.00451058	-0.00216209	0.311871	0.0491332	-0.0813806	0.0329838	-0.0243665
PREG5	0.180876	0.0975996	-0.294824	-0.0128859	-0.168062	0.12246	0.0646848	0.252605
PREG6	0.222628	-0.0423794	-0.0323457	0.344944	-0.0163959	-0.111236	-0.0440807	0.00743351
PREG7	0.201846	-0.0233152	-0.0756636	0.0824433	-0.275309	-0.173619	-0.0156208	0.0288045
PREG8	0.201093	0.0280099	0.120591	-0.060565	-0.256566	-0.187437	-0.0234847	-0.0176549
PREG9	0.165637	-0.0018381	0.178844	-0.262423	-0.128443	-0.160497	0.136648	0.0548024
PREG10	0.188457	0.040128	-0.0565107	-0.158829	-0.121057	-0.308821	-0.184254	-0.0429747
PREG11	0.0801328	0.0322846	0.00934901	-0.134813	0.281726	-0.186369	-0.620573	-0.233995
PREG12	0.155335	0.0872228	0.0881686	-0.202783	-0.131287	0.238126	-0.222115	-0.0121403
PREG13	0.135473	0.236591	-0.0641644	-0.213565	0.00114663	-0.0134513	0.0853365	0.214078
PREG14	0.187348	-0.0239281	-0.0191149	0.0368457	-0.241222	0.0524846	0.0392816	-0.31792
PREG15	0.145545	-0.0657295	0.0119306	0.00326593	-0.308216	0.356366	-0.161607	-0.373974
PREG16	0.197295	0.164632	-0.147114	-0.0857819	0.0317884	-0.107942	-0.0917348	0.162297
PREG17	0.207661	-0.0173061	0.11172	0.165635	-0.0779378	-0.0514788	0.149304	-0.0905358
PREG18	0.0945915	0.108227	-0.334467	-0.133237	0.2225	0.278349	0.0684338	-0.230128
PREG19	0.159949	0.113683	-0.398378	-0.00822496	-0.0673757	0.0848417	0.167684	-0.0590253
PREG20	0.208435	0.134934	-0.126545	-0.167974	-0.00355984	0.045821	-0.0489778	0.0703549
PREG21	0.17336	0.0723876	0.103164	-0.0540708	0.220148	-0.306508	0.0310765	0.0270793
PREG22	-0.0557192	0.343343	0.111939	0.295932	-0.0726129	0.138867	-0.294434	-0.00907224
PREG23	0.0360992	0.378853	0.145476	0.128734	-0.0156273	0.158506	-0.0731441	0.305216
PREG24	-0.0283285	0.363951	0.111742	0.293757	-0.00695081	0.11593	-0.143744	0.112273
PREG25	0.196668	-0.0598516	0.208821	-0.0975041	-0.25159	0.215418	1.54875E-05	-0.111738
PREG26	0.161233	0.0344699	0.100848	-0.0794021	0.277341	0.241766	0.0255704	0.10965
PREG27	-0.14649	0.350183	0.0897017	-0.0863846	-0.0519477	-0.04619	0.20685	-0.287083
PREG28	0.136684	0.0324535	0.0211723	-0.135206	0.249251	0.2991	0.0647546	-0.0121728
PREG29	0.207958	-0.0132654	0.0736808	0.140659	0.202594	-0.139697	0.18699	-0.178405
PREG30	0.139706	0.0285318	0.410773	-0.101497	0.144673	0.0838648	0.00999597	0.0635281
PREG31	-0.151839	0.386036	0.0167417	0.06954	-0.0314784	-0.158287	0.156812	-0.206215
PREG32	0.166835	-0.0707581	0.0397675	0.223016	0.243572	0.144026	0.300098	-0.0627569
PREG33	-0.143329	0.357552	0.0602481	-0.134225	-0.0365706	-0.126051	0.184284	-0.293441

El primer componente explica el 31.42% de la variabilidad total y las variables explicadas son:

componente 1	
2. No podía quitarme la tristeza	0.22011
3. Tenía dificultad para mantener mi mente en lo que estaba haciendo	0.211156
4. Me sentía deprimido	0.217712
6. Me sentía triste	0.222628
7. No podía seguir adelante	0.201846
8. Nada me hacía feliz	0.201093
17. Estaba a disgusto conmigo mismo(a)	0.207661
20. Era difícil concentrarme en las cosas importantes	0.208435
29. Me sentía solo(a)	0.207958
34. Sentí que iba a darme por vencido(a)	0.20906

El segundo componente explica el 7.76% de la variabilidad total y las variables explicadas son:

componente 2	
22. Sentí que era tan bueno(a) como otra gente	0.343343
23. Sentí que todo lo que hacia era con esfuerzo	0.378853
24. Me sentía esperanzado(a) hacia el futuro	0.363951
27. Me sentía feliz	0.350183
31. Disfruté de la vida	0.386036
33. Me divertí mucho	0.357552

El tercer componente explica el 4.75% de la variabilidad total y las variables explicadas son:

componente 3	
18. Perdí peso sin intentarlo	-0.334467
19. Me costaba mucho trabajo dormir	-0.398378
30. Las personas eran poco amigables	0.410773
35. Sentí que le desagradaba a la gente	0.382324

El cuarto componente explica el 4.30% de la variabilidad total y las variables explicadas son:

componente 4	
2. No podía quitarme la tristeza	0.348647
4. Me sentía deprimido	0.311871
6. Me sentía triste	0.344944
22. Sentí que era tan bueno(a) como otra gente	0.295932
24. Me sentía esperanzado(a) hacia el futuro	0.293757

El quinto componente explica el 3.87% de la variabilidad total y las variables explicadas son:

componente 5	
1. Tenía poco apetito	0.295933
7. No podía seguir adelante	-0.275309
8. Nada me hacía feliz	-0.256566
25. Pensé que mi vida ha sido un fracaso	-0.25159
26. Me sentía temeroso(a)	0.277341

El sexto componente explica el 3.77% de la variabilidad total y las variables explicadas son:

componente 6	
10. Había perdido interés en mis actividades diarias	-0.308821
15. Quería hacerme daño	0.356366
21. Me molestaba por cosas que usualmente no me molestan	-0.306508
28. Hablé menos de lo usual	0.2991

El séptimo componente explica el 3.33% de la variabilidad total y las variables explicadas son:

componente 7	
11. Dormía más de lo habitual	-0.620573
32. Tenía ataques de llanto	0.300098

El octavo componente explica el 3.20% de la variabilidad total y las variables explicadas son:

componente 8	
1. Tenía poco apetito	-0.291842
14. Sentía deseos de estar muerto	-0.31792
15. Quería hacerme daño	-0.373974
23. Sentí que todo lo que hacia era con esfuerzo	0.305216
33. Me divertí mucho	-0.293441

Diagrama de Dispersión

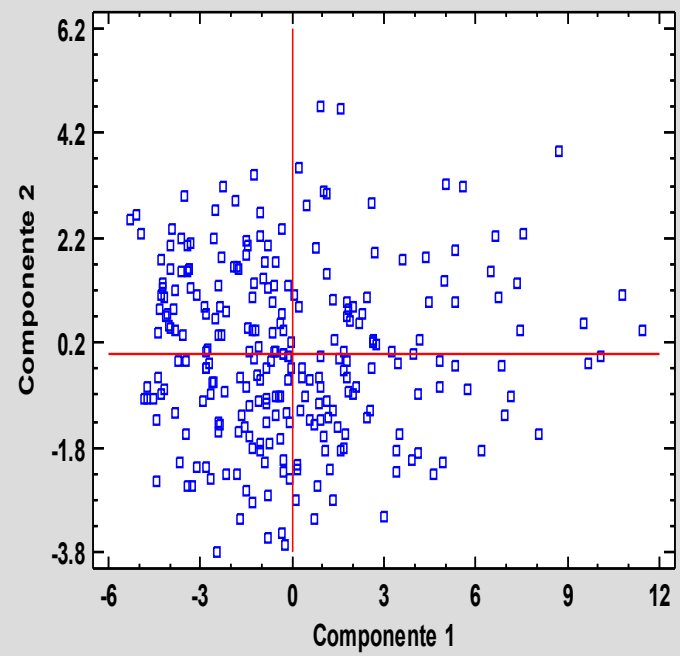
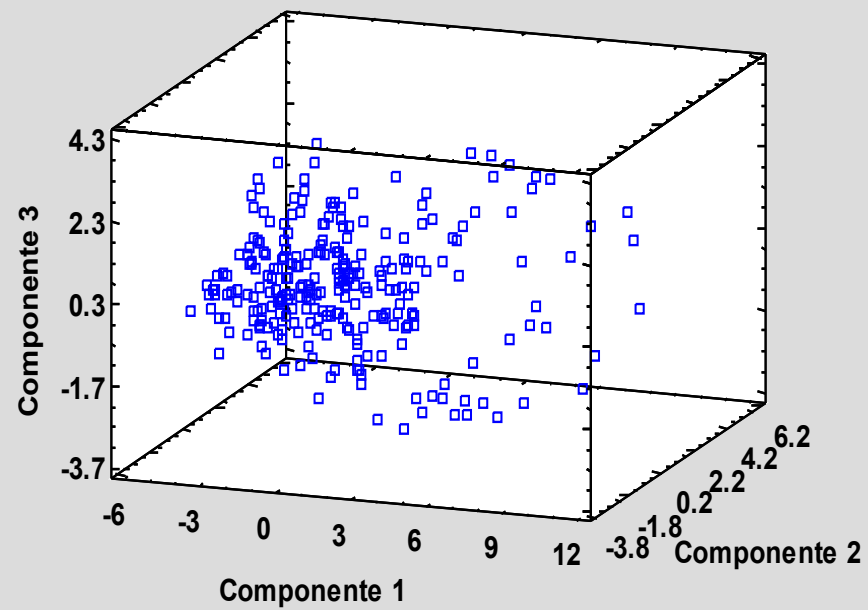
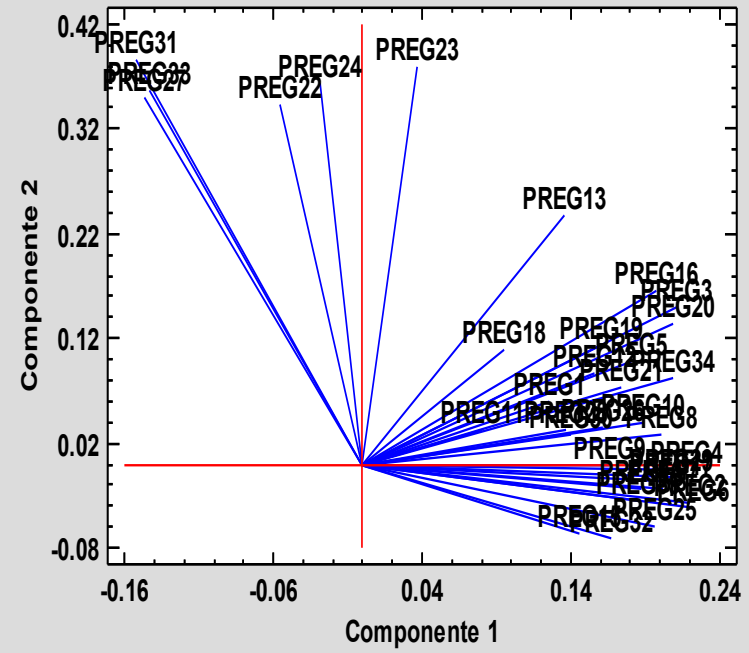


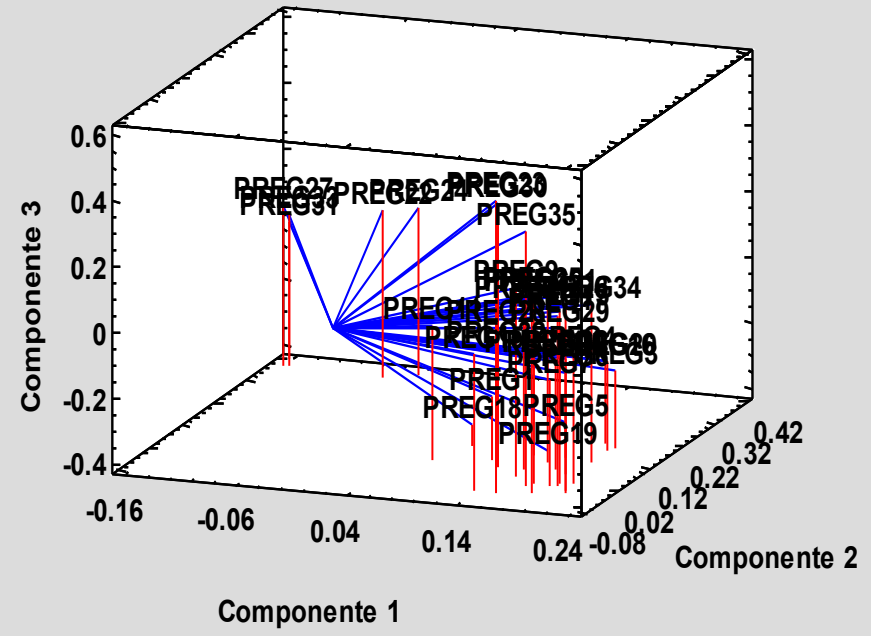
Diagrama de Dispersión



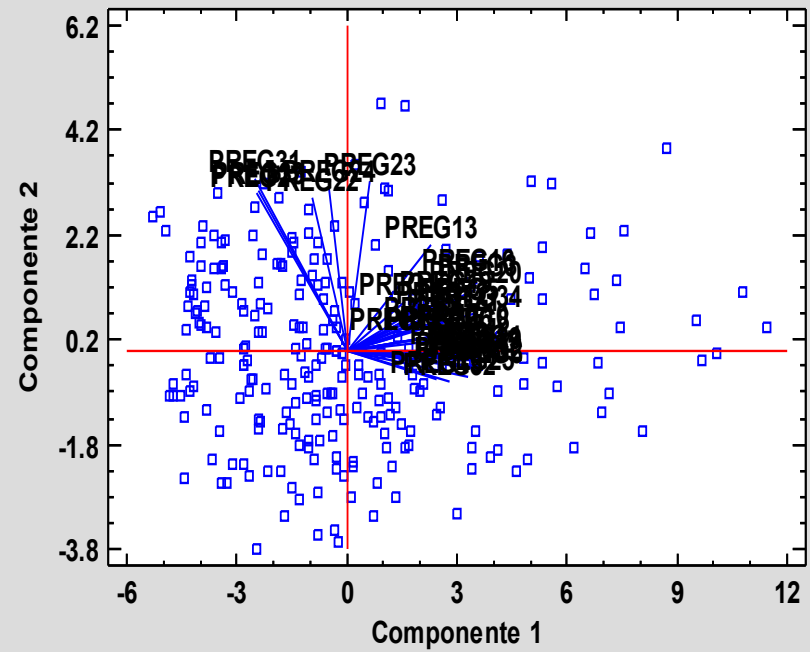
Gráfica de Pesos del Componente



Gráfica de Pesos del Componente



Bigráfica



Bigráfica

