

# ESTUDIO DE LAS CONDICIONES DE DESARROLLO DE UN CONJUNTO DE DISTRITOS DE LA REPÚBLICA DE PANAMÁ

JUAN DÍAZ\*

GONZALO CARRASCO†

*Recibido/Received: 26/05/05 — Aceptado/Accepted: 22/09/06*

---

## Resumen

A través de las componentes principales, se reflejan las ponderaciones de cada una de las variables, que a su vez, permitirán agrupar a distritos de la República de Panamá en grupos semejantes a las condiciones de desarrollo y así, determinar las similitudes y desigualdades.

**Palabras clave:** componente principal, valor propio, vector propio, matriz de correlación, matriz de varianza-covarianza, ortogonal, distrito, desarrollo.

## Abstract

Through the Principal Components the weights of each one of the variables are reflected and as well they will allow to group the districts of the Republic of Panama in similar conditions of development groups and thus for determining the similarities and inequalities.

**Keywords:** principal component, eigenvalue, eigenvector, correlation matrix, variance-covariance matrix, orthogonal, district, development.

**Mathematics Subject Classification:** 62H25

---

\*Escuela de Estadística, Universidad de Panamá.

†Centro de Investigación y Consultoría Estadística (CICE), Departamento de Estadística, Universidad de Panamá. E-mail: [gcarrasco27@hotmail.com](mailto:gcarrasco27@hotmail.com)

## 1 Introducción

El presente trabajo tiene como objetivo central, aplicar el análisis de las componentes principales (ACP) a un conjunto de datos procedentes de 66 distritos de la República de Panamá, con el fin de conocer las condiciones de desarrollo. Para eso, se utilizaron 17 variables que son consideradas factores relevantes en el análisis de desarrollo de los distritos del país.

A través del ACP, se agruparán los distritos en grupos semejantes a las condiciones de desarrollo para así, conocer las igualdades y diferencias existentes entre ellos.

En este artículo, se presentará en primer lugar, un breve resumen del ACP; en segundo lugar, se mostrará la población sujeta de estudio y las variables involucradas en el mismo y en tercer lugar, se brindará un análisis de los resultados.

## 2 Análisis en componentes principales

El análisis de las componentes principales (ACP), es uno de los métodos de análisis multivariado que permite explicar la estructuración de las variancias y covarianzas de un conjunto de variables correlacionadas mediante combinaciones lineales de las variables originales, llamadas componentes principales. Estas componentes principales no están correlacionadas entre sí y cada una maximiza su varianza.

El ACP simplifica o reduce los datos y facilita su interpretación. De esta manera, el conjunto original de  $p$  variables, se reducirá a  $q$  componentes ( $q < p$ ). Esta generación de las componentes principales, permitirá descubrir rasgos más interesantes para fines de interpretación o descriptivos, los cuales hubiera podido lograr analizando la totalidad de las variables originales.

Al efectuar un análisis de componentes, el investigador frecuentemente calcula el valor que toma cada componente en cada individuo. A estos valores se les llama puntaje componentes, que a su vez servirán como datos de entrada para la utilización de otras técnicas que acepten como variables de entrada a las componentes principales.

## 3 Estudio de los distritos de la República de Panamá

### 3.1 Población

La población objeto de estudio, está integrada por 66 distritos de la República, a los cuales se le midieron las características de las viviendas, condición de educación, tasa de mortalidad infantil y niveles de educación superior de sus habitantes.

Los distritos que fueron incluidos en este estudio fueron los mostrados en la tabla 1.

### 3.2 Procedencia de los datos

La información utilizada tiene como fuente de origen:

- Censos Nacionales

Provincia	Distritos
Bocas del Toro	Bocas del Toro, Changuinola, Chiriquí Grande
Coclé	Aguadulce, Antón, La Pintada, Natá, Olá, Penonomé
Colón	Colón, Chagras, Donoso, Portobelo, Santa Isabel, San Blas
Chiriquí	Alanje, Barú, Boquerón, Boquete, Bugaba, David, Dolega, Gualaca, Remedios, Renacimiento, San Félix, San Lorenzo, Tolé
Darién	Chepigana, Pinogana
Herrera	Chitré, Las Minas, Los Pozos, Ocú, Parita, Pesé, Santa María
Los Santos	Guararé, Las Tablas, Los Santos, Macaracas, Pedasí, Pocrí, Tonosí
Panamá	Arraiján, Balboa, Capira, Chame, Chepo, Chimán, La Chorrera, Panamá, San Carlos, San Miguelito, Taboga
Veraguas	Atalaya, Calobre, Caazas, La Mesa, Las Palmas, Montijo, Río de Jesús, San Francisco, Santa Fe, Santiago, Soná

Tabla 1: Provincias y distritos incluidos en el análisis.

- Estadísticas de salud, publicada por el Ministerio de Salud y de la Caja de Seguro Social (C.S.S.).

La información procedente del Censo de Población y Vivienda, está referida al año 1990, de ella se obtuvo la información sobre las variables de vivienda. La información de las variables de salud, procede en parte de las Estadísticas Vitales de la Contraloría General de la República y la otra parte, proviene del Departamento de Registros Médicos y Estadísticas del Ministerio de Salud y de la C.S.S.

Para procesar los datos, se utilizó el paquete Excel (Macro de Excel© XL STAT.).

### 3.3 Definición de variables

Las variables que se incluyeron en el análisis de componentes, son las que se describen a continuación:

**INCON** : Índice de condición de enseñanza primaria y secundaria (refleja el valor de aspectos tales como: personal docente, población en edad escolar que asiste a las escuelas).

**ING** : Mediana del ingreso familiar.

**TMI** : Tasa de mortalidad infantil.

**SLUZ** : Porcentaje de viviendas sin luz eléctrica.

**PISO** : Porcentaje de viviendas con pisos inadecuados.

**AGUA** : Porcentaje de viviendas con agua potable.

**TEL** : Porcentaje de viviendas con teléfono.

**TV** : Porcentaje de vivienda con televisión.

**SUNI** : Porcentaje de población sin educación universitaria.

**SINGRAD** : Porcentaje de población sin ningún grado escolar.

**PEA** : Porcentaje de personas económicamente activas, es decir, categorizadas como ocupadas en su mayor parte y desocupadas.

**URB** : Porcentaje de urbanismo en el distrito.

**PERVIV** : Promedio de personas por vivienda.

**CUARTVIV** : Promedio de cuartos por vivienda.

**DESN** : Porcentaje de población infantil desnutrida.

**TASACREC** : Tasa de crecimiento poblacional.

**MEDICO** : Porcentaje de médicos, según población.

## 4 Resultados

### 4.1 Correlaciones

Fueron incluidas en el análisis de componentes principales un total de 17 variables, las cuales se encuentran correlacionadas entre sí, lo que permite la realización del tipo de análisis anteriormente mencionado.

Una vez examinada la matriz de correlación, procedemos a efectuar el análisis de componentes principales a partir de ella, con el objeto de que todas las variables que intervienen en el estudio tengan el mismo grado de importancia.

### 4.2 Determinación del número de componentes principales

En la Tabla 3, se presentan los valores propios a partir de la matriz de correlación. Obsérvese que en este caso, la suma de los valores propios es de 17, el número total de variables consideradas.

El primer componente sintetiza el 50.83% de la variabilidad, el segundo el 18.98% y el tercero el 8.04% . Así los tres primeros componentes explican una alta proporción, es decir, 77.85% de la variabilidad total. Estos componentes pueden, por lo tanto, representar a las 17 variables originales sin pérdida importante de información.

Aplicando el criterio de selección de componentes principales de Kaiser, se podrán utilizar si es necesario, los tres primeros componentes, ya que sus valores propios correspondientes son mayores que el promedio, es decir, superiores a 1. El método gráfico representado por la gráfica 1, nos lleva a una misma conclusión, observándose un punto de cambio en el tercer componente.

	suni	singrad	pea	ing	tmi	urb	tv	sluz	tel
Suni	1.00	0.59	-0.46	-0.89	0.57	-0.80	-0.84	0.82	-0.90
Singrad	0.59	1.00	-0.62	-0.54	0.63	-0.36	-0.76	0.75	-0.42
Pea	-0.46	-0.62	1.00	0.35	-0.58	0.14	0.51	-0.51	0.34
Ing	-0.89	-0.54	0.35	1.00	-0.52	0.88	0.85	-0.82	0.86
Tmi	0.57	0.63	-0.58	-0.52	1.00	-0.37	-0.65	0.66	-0.48
Urb	-0.80	-0.36	0.14	0.88	-0.37	1.00	0.72	-0.68	0.80
Tv	-0.84	-0.76	0.51	0.85	-0.65	0.72	1.00	-0.98	0.77
Sluz	0.82	0.75	-0.51	-0.82	0.66	-0.68	-0.98	1.00	-0.75
Tel	-0.90	-0.42	0.34	0.86	-0.48	0.80	0.77	-0.75	1.00
Perviv	0.20	0.51	-0.53	-0.15	0.43	0.06	-0.36	0.39	-0.10
Cuartviv	-0.09	-0.26	0.14	0.13	-0.11	0.08	0.17	-0.19	0.01
Incon	0.22	-0.04	0.25	-0.42	0.02	-0.52	-0.19	0.14	-0.27
Piso	0.71	0.82	-0.48	-0.74	0.59	-0.60	-0.89	0.91	-0.61
Agua	-0.70	-0.82	0.65	0.65	-0.64	0.52	0.85	-0.86	0.57
Desn	-0.59	-0.14	0.05	0.62	-0.17	0.63	0.37	-0.32	0.60
Tasacrec	-0.05	0.32	-0.49	0.24	0.26	0.50	-0.03	0.07	0.14
Medicos	-0.76	-0.35	0.33	0.63	-0.43	0.61	0.57	-0.58	0.70

	perviv	cuartviv	incon	piso	agua	desn	tasacrec	medicos
Suni	0.20	-0.09	0.22	0.71	-0.70	-0.59	-0.05	-0.76
Singrad	0.51	-0.26	-0.04	0.82	-0.82	-0.14	0.32	-0.35
Pea	-0.53	0.14	0.25	-0.48	0.65	0.05	-0.49	0.33
Ing	-0.15	0.13	-0.42	-0.74	0.65	0.62	0.24	0.63
Tmi	0.43	-0.11	0.02	0.59	-0.64	-0.17	0.26	-0.43
Urb	0.06	0.08	-0.52	-0.60	0.52	0.63	0.50	0.61
Tv	-0.36	0.17	-0.19	-0.89	0.85	0.37	-0.03	0.57
Sluz	0.39	-0.19	0.14	0.91	-0.86	-0.32	0.07	-0.58
Tel	-0.10	0.01	-0.27	-0.61	0.57	0.60	0.14	0.70
Perviv	1.00	-0.15	-0.41	0.34	-0.48	0.16	0.42	-0.09
Cuartviv	-0.15	1.00	-0.18	-0.23	0.22	-0.04	0.05	-0.09
Incon	-0.41	-0.18	1.00	0.12	0.01	-0.33	-0.54	-0.03
Piso	0.34	-0.23	0.12	1.00	-0.81	-0.26	0.02	-0.48
Agua	-0.48	0.22	0.01	-0.81	1.00	0.22	-0.29	0.45
Desn	0.16	-0.04	-0.33	-0.26	0.22	1.00	0.28	0.56
Tasacrec	0.42	0.05	-0.54	0.02	-0.29	0.28	1.00	0.02
Medicos	-0.09	-0.09	-0.03	-0.48	0.45	0.56	0.02	1.00

Tabla 2: Matriz de correlación.

### 4.3 Vectores propios

Los coeficientes del primer vector señalado en la Tabla 4, indica que esta combinación lineal de todas las variables incluidas en el análisis, presenta ponderaciones con valores positivos relativamente elevados para TV (0.33), ING (0.31), TEL (0.29), AGUA (0.29), URB (0.26), MEDICOS (0.24), PEA (0.20), y con valores negativos para variables como SIUZ (-0.32), SUNI (-0.31), PISO (-0.30) y TMI (0.24). Es decir, un incremento de las variables con valores positivos, influye en una disminución de las variables con valores negativos.

### 4.4 Matriz de correlación de las variables originales con los componentes principales

La gráfica 2, representa el círculo de radio unitario que permite identificar la correlación de las variables originales con los dos primeros componentes. Aquellas variables que se acercan más al círculo, son las que más se correlacionan con las componentes y las que

	Valor propio	% Varianza	Varianza acumulada
1	8.64	51%	51%
2	3.22	19%	70%
3	1.36	8%	78%
4	0.70	4%	82%
5	0.594	4%	86%
6	0.505	3%	88%
Total	17		100%

Tabla 3: Valores propios.

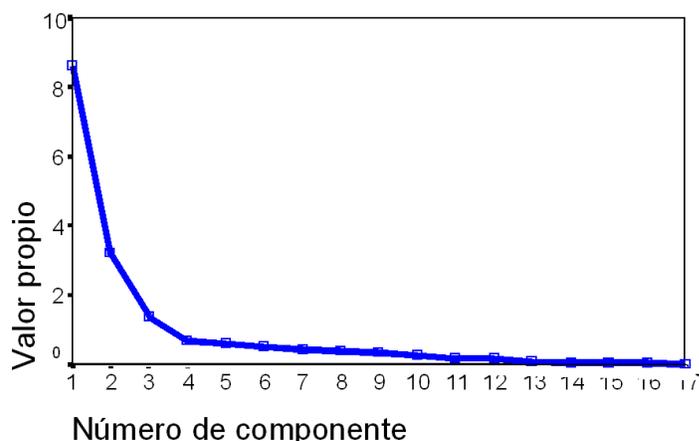


Figura 1: Gráfico de valores propios.

se agrupan cerca del centro, serán las que menos lo hacen. Es decir, las variables mejor explicadas se ubicarán cerca del círculo, o en el centro en caso contrario.

Se puede observar en la Tabla 5, que los componentes tienen correlaciones moderadas y altas con las variables que se indican seguidamente:

El primer componente está correlacionado con: SUNI, SINGRAD, PEA, INGR, TMI, URB, TV, SLUZ, TEL, PISO, AGUA y MEDICOS.

El segundo componente está dominado por las siguientes variables: PERVIV, INCON, TASACREC.

Para el tercer componente, éste se correlaciona sólo con la variable CUARTVIV.

#### 4.5 Interpretación de los componentes generados

Utilizando la Tabla 5, se observa que la contribución de las variables en el primer componente, revelan que los distritos con valores altos en este, estarán asociados a viviendas con características deseables, es decir, con servicios básicos de agua potable, luz eléctrica,

	Vector propio		
	1	2	3
suní	-0.31	-0.09	0.14
singrad	-0.26	0.24	-0.17
pea	0.20	-0.32	-0.09
ing	0.31	0.18	0.02
tmi	-0.24	0.17	0.02
urb	0.26	0.31	0.03
tv	0.33	-0.02	0.08
sluz	-0.32	0.05	-0.08
tel	0.29	0.17	-0.18
perviv	-0.12	0.38	-0.03
cuartviv	0.06	-0.06	0.66
incon	-0.07	-0.39	-0.37
piso	-0.30	0.07	-0.18
agua	0.29	-0.18	0.09
desn	0.17	0.30	-0.26
tasacrec	-0.01	0.45	0.22
medicos	0.24	0.10	-0.41

Tabla 4: Coordenadas de las variables originales en los primeros 3 vectores propios.

televisión, teléfono, pisos adecuados y un menor número de personas por vivienda. En lo relacionado a la educación, se caracterizan por bajos niveles de personas sin ningún grado de escolaridad aprobado y sin nivel universitario. En cuanto al poder adquisitivo de esta población, son relativamente altos debido a una elevada mediana de los ingresos familiares, asociada también con una alta proporción de población económicamente activa. Estos distritos su vez, están influenciados por altos niveles de urbanismo y mejores condiciones de salud, observado en una baja tasa de mortalidad infantil y un alto porcentaje de médicos según población.

Este primer componente se puede interpretar como un “ÍNDICE DE DESARROLLO SOCIAL”, considerando la naturaleza de las variables que lo integran. Como este componente sintetiza la mayor parte de la información y además, clasifica los distritos según las condiciones de desarrollo, utilizaremos únicamente este componente para agrupar los distritos.

#### 4.6 Puntajes componentes

En la tabla 6, presentamos los puntajes que se le asignan a cada distrito según el Índice de Desarrollo Social.

Los distritos con mejores condiciones de vida y mayor desarrollo social, son los que presentan puntajes positivos elevados en el primer componente, lo cual se puede constatar con la Tabla 6, donde vemos que el Distrito de Panamá presenta el mayor puntaje (8.6127), seguido del Distrito de San Miguelito (6.3315), el tercer lugar lo ocupa el Distrito de Chitré

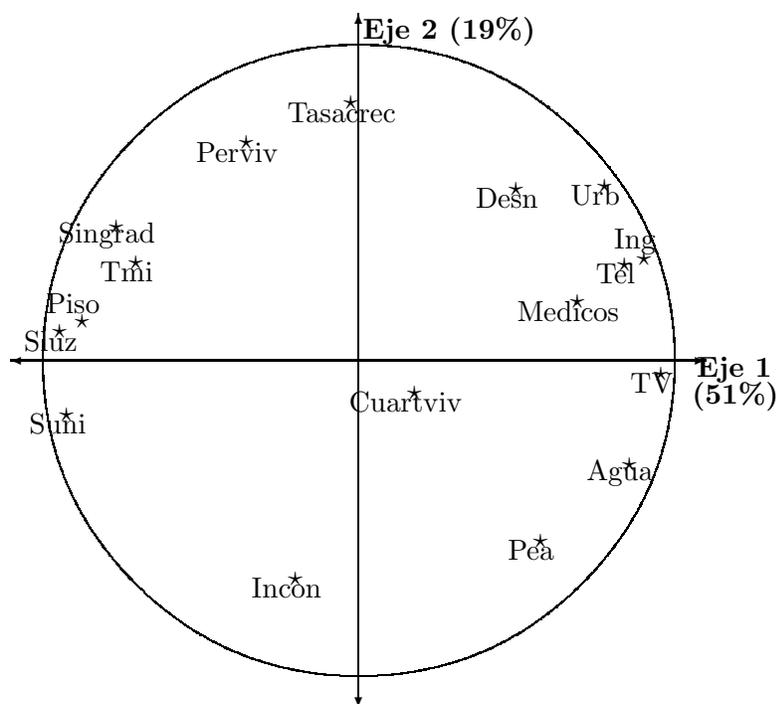


Figura 2: Círculo de correlación de las componentes 1 y 2.

(5.8929), seguido de David, Colón, Arraiján, La Chorrera y Aguadulce.

Como podemos observar, los distritos con estos puntajes se ubican en su mayor parte en la provincia de Panamá y la cabeceras de las provincias del interior del país, lo que permite un mejor desarrollo de estas regiones. Se encuentra un número importante de distritos que forman parte de la provincia de Panamá, lo cual muestra un mayor grado de desarrollo social en estas localidades cercanas al área metropolitana y que como Arraiján y Chorrera, están en un constante crecimiento poblacional y urbanístico debido a una construcción importante de complejos habitacionales.

Los distritos con valores negativos son los distritos con un menor desarrollo social. Podemos observar que estos distritos que comparten características de menor desarrollo en el país debido al grado de inaccesibilidad de estas regiones: Darién, Bocas del Toro, San Blas, el marcado número de población indígena y carencias de los servicios básicos. Dentro de éstos destacan: Chiriquí Grande, Tolé, Remedios, San Blas, San Lorenzo y Bocas del Toro; (ver tabla 6).

## 5 Conclusión

El primer componente sintetiza la mayor parte de la información y considerando la naturaleza de las variables que lo integran, permite agrupar a los distritos según las condiciones

Variable	Componente principal		
	1	2	3
Suni	-0.9228	-0.1705	0.1648
singrad	-0.7664	0.4241	-0.1978
Pea	0.5762	-0.5711	-0.1008
ing	0.9043	0.3213	0.0217
tmi	-0.7043	0.3089	0.0179
urb	0.7789	0.5527	0.0343
tv	0.9581	-0.0430	0.0927
sluz	-0.9452	0.0914	-0.0929
tel	0.8432	0.3031	-0.2091
perviv	-0.3550	0.6894	-0.0298
cuartviv	0.1778	-0.1035	0.7704
incon	-0.1983	-0.6922	-0.4362
piso	-0.8739	0.1242	-0.2128
agua	0.8575	-0.3284	0.1076
desn	0.4993	0.5435	-0.3058
tasacrec	-0.0233	0.8162	0.2614
medicos	0.6939	0.1861	-0.4741

Tabla 5: Matriz de correlación de las variables originales con los primeros componentes.

de desarrollo.

Podemos señalar que los distritos con mejores condiciones de vida y mayor desarrollo social, se ubican sobre todo en las áreas urbanas del país, obteniendo el mayor puntaje el Distrito de Panamá (8.6127), seguido del Distrito de San Miguelito (6.3315), el tercer lugar lo ocupa el Distrito de Chitré (5.8929), seguido de David, Colón, Arraiján, La Chorrera y Aguadulce.

Contraste a lo anterior, se observa que los distritos con condiciones de vida desfavorables para sus habitantes se ubican en las áreas rurales e indígenas. Dentro de éstos destacan: Chiriquí Grande, Tolé, Remedios, San Blas, San Lorenzo, Cañazas y Bocas de del Toro.

Se debe exhortar a las instituciones de nuestro país encargadas de la recolección y publicación de cifras estadísticas, al uso del ACP, ya que permite una evaluación completa de la información a través de un número de componentes principales que es menor al número de variables originales.

## Referencias

- [1] Anderson, R.E.; Black, W.; Hair, J.E.; Thatam, R.L. (1999) *Análisis Multivariante*. Prentice-Hall, Madrid.

Distrito	Puntaje	Distrito	Puntaje	Distrito	Puntaje
1. Panamá	8.6127	23. Natá	0.7977	45. Montijo	-2.153
2. San Miguelito	6.3315	24. Pesé	0.7119	46. Río de Jesús	-2.2144
3. Chitré	5.8929	25. Pocrí	0.6841	47. Calobre	-2.2612
4. David	5.3749	26. San Carlos	0.4767	48. San Félix	-2.3297
5. Colón	5.2044	27. Portobelo	0.3391	49. La Mesa	-2.3638
6. Arraiján	4.7107	28. Antón	0.2298	50. Los Pozos	-2.3812
7. La Chorrera	4.4787	29. Penonomé	0.2203	51. Chimán	-2.4442
8. Aguadulce	4.1641	30. Balboa	0.0563	52. Chepigana	-2.4861
9. Santiago	3.8292	31. Chepo	0.0129	53. Pinogana	-2.5151
10. Las Tablas	3.7419	32. Alanje	-0.1456	54. Donoso	-2.5125
11. Los Santos	3.1492	33. Atalaya	-0.1577	55. San Francisco	-2.6358
12. Taboga	2.518	34. Capiro	-0.1935	56. Remedios	-3.078
13. Boquete	2.4548	35. Boquerón	-0.263	57. San Blas	-3.1334
14. Changuinola	2.104	36. Macaracas	-0.4135	58. San Lorenzo	-3.1846
15. Guararé	1.8228	37. Ocú	-0.8426	59. Bocas del Toro	-3.392
16. Barú	1.7475	38. Gualaca	-0.9741	60. Santa Fe	-3.4172
17. Bugaba	1.6716	39. Tonosí	-1.0814	61. Las Minas	-3.4338
18. Chame	1.6435	40. Renacimiento	-1.2038	62. Olá	-3.4703
19. Dolega	1.478	41. Santa Isabel	-1.2166	63. Las Palma	-3.6765
20. Santa María	1.3665	42. Soná	-1.3077	64. Cañazas	-3.8146
21. Parita	1.2729	43. La Pintada	-1.8164	65. Tolé	-4.3431
22. Pedasí	0.988	44. Chagres	-1.8709	66. Chiriquí Grande	-5.3594

Tabla 6: Puntaje componente “Índice de Desarrollo Social”.

- [2] Anderson, T.W. (1958) *An Introduction Multivariate Analysis*. John Wiley and Sons, New York.
- [3] Carrasco, G. (1996) *Análisis de las Componentes Principales y sus Aplicaciones*. Tesis de la Universidad de Panamá, Panamá.
- [4] Carrasco, G.; Mejía, A. (1997) “Valores propios de la matriz de varianza covarianza muestral y sus distribuciones”, *Scientia* **13**(1): 79–85.
- [5] Johnson, D.E. (2000) *Métodos Multivariado aplicado al Análisis de Datos*. Thompson.
- [6] Johnson, R.; Wichern, D. (1982) *Applied Multivariate Statistical Analysis*. Prentice Hall, New Jersey.
- [7] Jolliffe, I.T. (1986) *Principal Component Analysis*. Springer-Verlag, New York.
- [8] Mardia, K.V.; Kent, J.T.; Bibby, J.M. (1979) *Multivariate Analysis*. Academic-Press, London.
- [9] Morrison, D.F. (1976) *Multivariate Statistical Methods*. McGraw-Hill, New York.
- [10] Pla, L.E. (1986) *Análisis Multivariado. Método de Componentes Principales*. O.E.A., Washington.