

COMPONENTES PRINCIPALES EN LA DETERMINACIÓN DE ESTACIONES CON PATRONES HOMOGÉNEOS DE TEMPERATURA EN EL CHOCÓ.

Main Components in the Determination of Stations with Temperature Homogeneous Patterns in the Chocó

RESUMEN

Este artículo presenta una aplicación del Análisis de Componentes principales para determinar estaciones con patrones homogéneos de temperatura en el Departamento del Chocó (Colombia), Permitiendo determinar dos grupos de estaciones con patrones homogéneos. Un grupo que presenta poca variación en el comportamiento de la temperatura, (Región Sur Oriente, Norte, Centro Occidente, Centro, Centro Oriente y Sur), y otra con variaciones particulares, (Zona Nor Oriente.). El uso del Análisis de Componentes Principales permitió evidenciar que con la determinación de estaciones con patrones homogéneos de temperatura, se descubre estadísticamente la Zona más calurosa del departamento.

PALABRAS CLAVES: Análisis factorial de Componentes principales, Estaciones Meteorológicas, Patrones de Temperatura.

ABSTRACT

This paper presents an application of principal component analysis to identify seasons, with homogeneous temperature patterns in the province of Chocó (Colombia), allowing defining two groups of stations with homogeneous patterns. A group with little variation in the behavior of the temperature, (South East Region, North, Central West, Central, East Central and South), and another with different variations (North East Zone.). The use of principal component analysis allowed evidencing that the determination of homogeneous stations with temperature patterns, we discover statistically the warmest part of the department.

KEYWORDS: Principal components factor analysis, weather stations, temperature patterns.

1. INTRODUCCIÓN

1.1 ANÁLISIS DE COMPONENTES PRINCIPALES.^[1]

El análisis de componentes principales, es una técnica factorial que consiste en resumir la información contenida en una matriz de datos con V variables, usando un puente de reducción para encontrar nuevos ejes que representen todas las variables. En el análisis de componentes principales, las relaciones entre las variables se estudian a partir de correlaciones y la relación entre individuos a partir de la distancia (pagés, J(1996) El Análisis de Componentes Principales trata de hallar componentes (factores) que sucesivamente expliquen la mayor parte de la varianza total.

El Análisis de Componentes Principales busca hallar combinaciones lineales de las variables originales que expliquen la mayor parte de la variación total; en este, el primer factor o componente sería aquel que explica una mayor parte de la varianza total, el segundo factor sería aquel que explica la mayor parte de la varianza restante, es decir, de la que no explicaba el primero y así sucesivamente.

1.1.3. PASOS EN EL ANÁLISIS FACTORIAL DE COMPONENTES PRINCIPALES

Los pasos que se suelen seguir en el Análisis Factorial son:

- 1- Calcular la matriz de correlaciones entre todas las variables (Matriz R).
- 2- Extracción de los factores necesarios para representar los datos.

JORGE ANDRÉS URRUTIA

MSc. Investigación de Operaciones y estadística
Profesor Auxiliar
Universidad Tecnológica de Pereira
jurrutia@utp.edu.co

REINER PALOMINO LEMUS

M. Sc. En Meteorología
Profesor Investigador
Universidad Tecnológica del Chocó
reiner@utch.edu.co

- 3- Rotación de los factores con objeto de facilitar su interpretación. Representación gráfica.

1.12. EXAMEN DE LA MATRIZ DE CORRELACIONES^[2]

Una vez que se dispone de esta matriz concierne examinarla para comprobar si sus características son adecuadas para realizar un Análisis Factorial. Uno de los requisitos que deben cumplirse para que el Análisis Factorial tenga sentido es que las variables estén altamente correlacionadas.

Se puede comprobar el grado de asociación entre las variables son:

- *El determinante de la matriz de correlaciones:* un determinante muy bajo indicará altas intercorrelaciones entre las variables, pero no debe ser cero (matriz no singular), pues esto indicaría que algunas de las variables son linealmente dependientes y no se podrían realizar ciertos cálculos necesarios en el Análisis Factorial).

- *Test de Esfericidad de Bartlett:* Comprueba que la matriz de correlaciones se ajuste a la matriz identidad (**I**), es decir ausencia de correlación significativa entre las variables. Esto significa que la nube de puntos se ajustara a una esfera perfecta, expresando así la hipótesis nula por: $H_0: \mathbf{R} = \mathbf{I}$, es decir, que el determinante de la matriz de correlaciones es 1. $H_0: |\mathbf{R}| = 1$. La formula correspondiente asume la siguiente expresión:

$$\chi^2 = - \left[n-1 - \frac{1}{6} * (2*v+5) \right] * \ln |R| \tag{1.1}$$

Donde...

- n =tamaño muestral.
- v =número de variables.
- ln=logaritmo neperiano.
- R =matriz de correlaciones.

Si se acepta la hipótesis nula ($p>0.05$) significa que las variables no están intercorrelacionadas y por tanto no tiene mucho sentido llevar a cabo un Análisis Factorial.

Es muy útil cuando el tamaño muestral es pequeño.

- *Índice KMO de Kaiser-Meyer-Olkin:*

$$KMO = \frac{\sum_{i \neq j} \sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} \sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} \sum_{i \neq j} a_{ij}^2} \tag{1.2}$$

Donde...

- r_{ij} = correlación simple.
- a_{ij} = correlación parcial.

Valores bajos del índice KMO desaconsejan la utilización de Análisis Factorial. Como baremo para interpretar el índice KMO podría tomarse según Kaiser:

- 1 >= KMO >= 0.9 muy bueno
- 0.9 >= KMO >= 0.8 meritorio
- 0.8 >= KMO >= 0.7 mediano
- 0.7 >= KMO >= 0.6 mediocre
- 0.6 >= KMO > 0.5 bajo
- KMO <= 0.5 inaceptable

1.1.4. MATRIZ FACTORIAL DE COMPONENTES

A partir de una matriz de correlaciones, el Análisis Factorial extrae otra matriz que reproduce la primera de forma más sencilla. Esta nueva matriz se denomina matriz factorial y adopta la siguiente forma:

	1	2
1	P11	P21
2	P12	P22
3	P13	P23

Cada columna es un factor y hay tantas filas como variables originales. Los elementos P_{ij} pueden interpretarse como índices de correlación entre el factor i y la variable j , aunque estrictamente sólo son correlaciones cuando los factores no están correlacionados entre sí, es decir, son ortogonales. Estos coeficientes reciben el nombre de pesos, cargas, ponderaciones o saturaciones factoriales. Los pesos factoriales indican el peso de cada variable en cada factor.

1.1.5. EIGEN VALUES (VALORES PROPIOS)

Los Eigen valúes (λ), indica la cantidad total de varianza que explica ese factor para las variables consideradas como grupo.

Las cargas factoriales pueden tener como valor máximo 1, por tanto el valor máximo que puede alcanzar el valor propio es igual al número de variables.

$$\begin{aligned} \lambda_1 &= P_{11}^2 + P_{12}^2 + \dots + P_{1j}^2 \\ \lambda_2 &= P_{21}^2 + P_{22}^2 + \dots + P_{2j}^2 \end{aligned} \tag{1.3}$$

$$\frac{\lambda_1}{I^2} \text{ Varianza explicada por el primer Factor}$$

$$\frac{\lambda_2}{I^2} \text{ Varianza explicada por el segundo Factor.}$$

1.2. VARIABLES METEOROLÓGICAS^[3].

Las variables Meteorológicas, son variables físicas utilizadas para estudiar el tiempo atmosférico y su evolución, algunas de ellas son: La Precipitación, Humedad, Temperatura, ect para nuestro caso de aplicación, nuestra variable será la temperatura.

1.2. TEMPERATURA

La temperatura se entiende como una medida de energía. La temperatura está relacionada con la **energía calorífica de los rayos solares** y es importante porque determina la formación de las nubes, afecta los valores de **humedad atmosférica** o cantidad de vapor de agua que se encuentra en el aire, e influye en la **presión atmosférica**, es decir, la fuerza que ejerce el peso del aire sobre la superficie terrestre.

Las unidades de temperatura son °C, °F, °K, °Rankine, °Reamur, la conversión más común es de °C a °F. La temperatura hace que el aire se vuelva menos denso, produciéndose una menor presión. La presión atmosférica junto con la temperatura hace posible la formación de viento el cual va a presentar un movimiento desde las zonas de mayores presiones a las de menores presiones. Los efectos de las variaciones de estos patrones se ven a menudo reflejados en los ecosistemas cambiando las condiciones de vida de los seres vivos. Algunos efectos producidos por estas variaciones has sido estudiado por Poveda, en Colombia y por el Francisco Soley en costa rica, Este último utilizando métodos multivariados de descomposición de varianza como lo son: descomposición de valores singulares, componentes ortogonales, y correlaciones canónicas.

2. DESARROLLO

Las series temperatura, de las 9 estaciones meteorológicas con funcionamiento en el departamento del Chocó utilizadas ara nuestro estudio, fueron suministrada por el IDEAM. Las series están desde 1983 hasta 1990. Para el desarrollo de este estudio, se realizó con esta fue lo siguiente:

Lo primero que se realizó fue depurar aquellas estaciones con más del 10% de datos faltantes, luego se organizaron las estaciones con mayor cantidad de registros en un orden de mayor a menor, teniendo como referencia, el años de instalación y funcionamiento.

Para calcular los datos faltantes en las estaciones, se tomó un tramo de 5 años en donde todas las estaciones presentaban registros continuos y se realizaron Correlaciones cruzadas, con el fin de establecer modelos de regresión ajustados que permitieran encontrar los datos faltantes en cada una de las estaciones.

Una vez realizado este procedimiento con cada una de las estaciones que presentaron ausencia de datos menores al 10%, se aplicaron pruebas de confiabilidad orientadas a la verificación y ajuste de los datos, aplicando el método doble masa Recomendado por (OMM, 1983).

2.1 MATRIZ DE DATOS FORMADA PARA EL ESTUDIO.

La Tabla 1 presenta un resumen de los datos estandarizados de las estaciones con patrones de temperatura usados para nuestro estudio.

Años	Meses	EST1	EST2	EST3	EST4	EST5	EST6
1983	E	1,29	1,96	1,53	1,00	-0,86	1,64
1983	F	-1,07	0,05	1,07	0,22	0,15	1,92
1983	M	1,50	2,00	1,19	0,46	0,35	2,23
1983	A	1,30	2,21	0,60	1,34	2,22	2,16
1983	M	1,94	2,08	0,77	0,51	1,67	2,16
1983	J	1,70	1,47	0,77	0,98	1,50	2,12
1983	J	1,50	1,31	0,55	1,16	1,74	2,16
1983	A	1,15	1,26	-0,24	1,03	1,47	1,87
1983	S	0,42	0,33	-1,19	0,59	0,38	0,20
1983	O	-0,32	0,86	-0,75	-0,53	-0,54	-0,35
1983	N	0,86	0,80	-0,14	0,71	0,92	0,14
1983	D	0,06	0,49	-1,24	0,31	-0,64	-0,89
1984	E	-0,37	-0,55	-0,40	-1,44	-0,03	-0,85

TABLA 1 MATRIZ DATOPS ESTANDARIZADOS DE TEMPERATURA

2.2 DERARROLLO DEL ANÁLISIS DE COMPONENTES PRINCIPALES

Con el propósito de obtener grupos de estaciones con comportamientos homogéneos temperatura, que permitieran describir el comportamiento de la misma en el departamento del Chocó, se realizó un análisis de componentes principales, con el paquete estadístico SPSS 15; los resultados y análisis se presenta a continuación.

2.2.1 RESULTADOS Y ANÁLISIS

Soportado en los principios del análisis factorial de componentes principales, Al analizar la matriz de correlaciones, se puede observar que en esta, el valor del determinante de la matriz es: (Determinante = 3.07E-

005) lo que nos indica que existen altas inter correlaciones entre las variables y nos garantiza que las variables (Estaciones) son linealmente independientes.

En la Tabla 2 de prueba de **Bartlett** o la medida de adecuación muestral **KMO**, observamos un valor de 0.904 que como se indicó en la sección 1, se considera meritorio la utilización de la técnica de reducción ACP, lo que significa que las variables si se adecuan representativamente a los factores encontrados por el ACP.

KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		904
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	619.805
	gl	36
	Sig.	.000

TABLA 2 PRUEBA DE BARTLETT PARA ESTACIONES CON PRECIPITACIÓN

Al inspeccionar la matriz de la varianza total explicada de la tabla 3, para las estaciones con patrones de precipitación, se evidencia las dos primeras componentes, explican el 74% de la varianza total, siendo la primera componente la que mayor cantidad de varianza explicada con un total de **62,681%**, y la segunda el **12.292%**. La tercera el **7.187%** y el resto menos del **5%**.

Varianza total explicada

Componente	Autovaleores iniciales		Suma de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	5.641	62.681	5.641	62.681	62.681	5.395	59.944	59.944
2	1.106	12.292	1.106	12.292	74.974	1.353	15.030	74.974
3	.647	7.187			82.160			
4	.409	4.550			86.710			
5	.351	3.900			90.610			
6	.287	3.191			93.802			
7	.227	2.520			96.321			
8	.171	1.904			98.226			
9	.160	1.774			100.000			

TABLA 3 VARIANZA TOTAL EXPLICADA

Estudiando la matriz rotada de la tabla 4, se observa claramente, que se han generado dos componentes, mostrando que en consecuencia tenemos dos grupos de estaciones con patrones similares en el comportamiento de la temperatura, que nos permitirá entender mejor el comportamiento de la misma en el Departamento.

Matriz de componentes rotados(a)

	Componente	
	1	2
ISTMINA	.920	.065
LLORÓ	.856	.198
APTOCARA	.854	.039
SANISIDRO	.837	.241
APTOCOND	.823	-.111
TERESITA	.804	.245
SANJIPAL	.782	.127
SAUTATA	.673	.497
PANAMER	.026	.956

TABLA 4 MATRIZ DE COMPONENTES ROTADOS

Los grupos quedan conformados de la siguiente manera:

Grupo 1 = (Istmina, Lloró, Aeropuerto del Caraño, San Isidro, Aeropuerto de Condoto, Teresita, San José del Palmar y Sautata).

Grupo 2 = (Panaerica)

La figura 1 de componentes, nos ilustra mejor esta distribución:

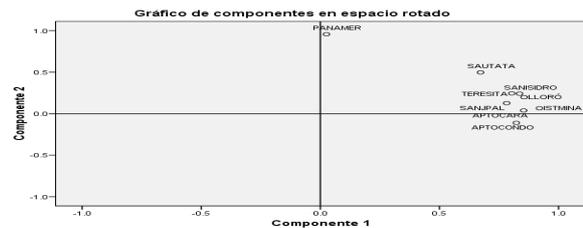


FIGURA 1 COMPONENTES PRINCIPALES

Gracias a la capacidad de extracción del método factorial de componentes principales, se puede describir el comportamiento de la temperatura del Departamento a partir de los dos grupos generados.

A continuación se presente una descripción del comportamiento típico de la temperatura para los dos grupos generados por las componentes.

En la Figura 2.1 se presenta el ciclo anual de la Temperatura en (°C) para estaciones que definieron el grupo1. Este grupo, no presenta picos definidos de temperatura, sin embargo, se observa una leve alza en el mes de *Diciembre* prolongada hasta finales del mes de *Marzo*, donde tiende a bajar. En términos generales, se observa que la Temperatura, tiende a permanecer constante, en todas las estaciones que conforman este grupo, quien está conformado por tres estaciones, pertenecientes a los municipios de (Lloró, Condoto, Istmina, Quibdó, Bahía Solano, San José del Palmar, Rio Sucio y Medio Atrato); las estaciones conforman la Región Sur Oriente, Norte, Centro Occidente, Centro, Centro Oriente y Sur.

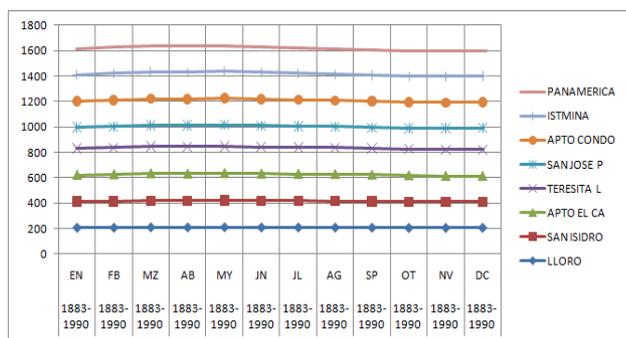


FIGURA 2 VALORES MEDIOS MENSUALES DE TEMPERATURA (°C) DE LA PRIMERA COMPONENTE 8 AÑOS

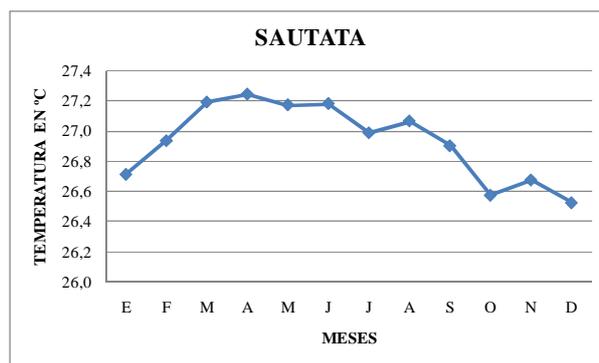


FIGURA 3 VALORES MEDIOS MENSUALES DE TEMPERATURA MEDIA DEL IRE EN (°C) DE LA SEGUNDA COMPONENTE 8 AÑOS

Estadísticos descriptivos

	N	Mínimo	Máximo	Media	Desv. típ.
LLORO	96	25.00	27.00	25.8260	.65038
SAN ISIDRO	96	25.00	28.00	26.0938	.59979
CARAÑO	96	20.00	28.00	26.3750	1.07850
TERESITA	96	25.00	28.00	26.3542	.56156
SAN JOSÉ DEL PALMAR	96	19.00	26.00	20.8021	1.12034
CONDOTO	96	24.00	28.00	25.9063	.66615
ISTMINA	96	24.70	28.00	26.0031	.63086
PANAMERICANA	96	23.00	26.00	25.3771	.66799
N válido (según lista)	96				

TABLA 5 ESTADÍSTICOS DESCRIPTIVOS

Los estadísticos descriptivos del grupo uno, presentado en la tabla 5, (Región Sur Oriente, Norte, Centro Occidente, Centro, Centro Oriente y Sur.), para cada una de las estaciones pertenecientes a él, muestra este grupo presenta valores promedios de temperatura en un intervalo de [25.37 °C. a 26.37°C]; valores máximos de temperatura en intervalo de [26°C a 28°C], y valores mínimos de Temperatura de [23°C a 25°C].

Este grupo, define unas regiones bastantes calurosas del departamento del Chocó.

En la Figura 2. se presenta el ciclo anual de la Temperatura en (°C) para estación que define el grupo 2. Este grupo Presenta valores máximos de temperatura bien definido en los meses de *Abril, Junio, Agosto y Noviembre*. Siendo el pico más alto en el mes de *Abril*. Un mínimo bien definido, se presenta en el mes de *Diciembre*, por prolongarse hasta mediados del mes de *Enero*, en donde se presenta un alza bastante considerable. Este grupo lo define la estación de Astuta pertenecientes al municipio de (Río Sucio); la estación conforma la Zona Nor Oriente del departamento.

Estadísticos descriptivos

	N	Mínimo	Máximo	Media	Desv. típ.
SAUTATA	96	26.00	29.00	26.9583	.59677
N válido (según lista)	96				

TABLA 6 ESTADÍSTICOS DESCRIPTIVO PARA CADA UNA DE LAS ESTACIONES DEL GRUPO

Los estadísticos descriptivos del grupo dos, presentado en la tabla 8.28, (Zona Nor Oriente.), para la estación que la define, muestra que esta zona presenta valores promedios de temperatura de [26°C]; valores máximos de temperatura de [29°C], y valores mínimos de Temperatura de [26°C].

Este grupo, define unas la zona más calurosa del Departamento del Chocó.

3. CONCLUSIONES Y RECOMENDACIONES

En esta investigación se ha analizado el comportamiento de los patrones de la temperatura en el Departamento del Chocó desde la perspectiva del análisis factorial de componentes principales. Se ha comprobado que el uso de la técnica es una herramienta eficiente para explicar y describir el comportamiento variacional de estos patrones meteorológicos.

El Análisis de Componentes Principales, aplicado a la determinación de estaciones con patrones de temperatura homogéneos en el departamento del Chocó, Permite establecer que existen dos tipos de comportamiento variacional de la temperatura en el Departamento, el primero lo determina la Región Sur Oriente, Norte, Centro Occidente, Centro, Centro Oriente y Sur, y el segundo la Zona Nor Oriente del departamento, lo que permite dar razón del porqué de la diversidad biológica en esta zona del país, y la vulnerabilidad de algunas regiones a inundaciones, sequías, y la proliferación de enfermedades como la malaria y el dengue.

Los resultados de este trabajo, se convertirán en una herramienta importante para la planeación a la prevención de desastres naturales, así como planes de prevención en salud, puesto que ha permitido establecer las diferentes condiciones de temperaturas que existen en el departamentos, punto de partida para delimitar las zonas con condiciones climáticas que son mas vulnerables a la propagación de enfermedades tropicales y desastres naturales.

4. BIBLIOGRAFÍA

[1] Peña, D. 2002. Análisis de Datos Multivariantes. Madrid: McGraw Hills/Interamericana de España.

[2] PÉREZ LÓPEZ, César, Técnicas de Análisis Multivariante de Datos, Aplicaciones con SPSS, Madrid, Universidad Complutense de Madrid."

[3] JORGE URRUTIA. Tesis de Maestría, Aplicación de Análisis Multivariado en la Determinación de la incidencia de variables Macroclimáticas en Patrones de Precipitación y Temperatura 2010.

[4] FRANCISCO SOLEY, Aplicación de Análisis Multivariado al Campo de Anomalías de la Precipitación en Centro America.