

# **PERFIL SOCIOECONÓMICO DE LOS ELECTORES UNIVERSITARIOS DE LA CIUDAD DE GUAYAQUIL MEDIANTE COMPONENTES PRINCIPALES NO LINEALES**

**Baque Bustamante Wilmer <sup>1</sup>, Ramirez Figueroa John <sup>2</sup>**

**1 Ingeniero en Estadística e Informática**

**2 M.Sc en Estadística, profesor de la ESPOL desde 1996, pertenece al centro de investigaciones estadísticas ICM-ESPOL**

## **RESUMEN.**

En el presente artículo se realiza un estudio de las condiciones socioeconómicas de los individuos, tales como, los ingresos, nivel educativo, calidad de los servicios básicos, etc., inciden en la forma de cómo estos eligen a sus dignidades. En el país existen pocos estudios que expliquen el comportamiento electoral de los ciudadanos en base a sus múltiples características, consideradas simultáneamente.

En el presente trabajo, se realizará una investigación del comportamiento del elector universitario guayaquileño, en la cual se identificarán un perfiles o patrones socioeconómicos de los estudiantes universitarios. Uno de los objetivos de este trabajo es, precisamente, analizar las particularidades de dichos perfiles. A partir de un conjunto de variables que describe a los electores, se identificará patrones de comportamiento, homogeneidad y heterogeneidad.

En el estudio, se hace una descripción y análisis de naturaleza estadística de la información obtenida a través de encuestas en las principales universidades de la ciudad. El procesamiento de los datos se ha hecho el módulo de Análisis de Componentes Principales no Lineales o Categóricas del software estadístico SPSS 10.

La presente tesis contará con el respectivo Análisis Univariado, Análisis Multivariado, las Conclusiones y Recomendaciones.

## **INTRODUCCIÓN**

En el análisis multivariado determinaremos el perfil para los electores universitarios de la ciudad de Guayaquil. Cuando se utilizan variables cualitativas y cuantitativas el análisis de perfiles coincide con el Análisis de Componentes Principales no Lineales o Categóricas.

Al aplicar el Análisis de Componentes No Lineales o Categóricas, se observarán grupos compuestos por variables, en donde se analizará si estos grupos se correlacionan unos con otros, los cuales determinan el perfil que se busca. Dicho análisis parte de  $p$  variables o características observables y dándole a cada variable una nueva cuantificación debido a la heterogeneidad de las escalas.

## **CONTENIDO**

### **1. Breve reseña política en el Ecuador**

Existiendo en el país más de 5 partidos políticos se puede emitir un criterio de los últimos 20 años (1978 -1998). En el cual nos dice que existen cuatro partidos políticos con mayor aceptación de los cuales se particiona totalmente entre regiones marcadas, es decir que existen dos partidos políticos de la sierra siendo estos la ID y la DP y dos de la costa el PRE y el PSC. El partido que ha conseguido escaños con mayor equidad regional ha sido el Partido Social Cristiano, toda vez que presenta una diferencia del 24 por ciento entre las dos regiones más grandes (la menor diferencia de los cuatro partidos analizados). Por el contrario, el partido con mayor concentración de apoyos en una sola región ha sido el Partido Roldosista Ecuatoriano, siendo además el que presenta mayor diferenciación entre las dos regiones (48 por ciento). En tanto, los partidos que tienen mayor votación en la sierra, tienen bajos apoyos en la costa (13 y 16 por ciento) y una diferencia regional de cerca del 43 por ciento.

En la Costa se da una mayor concentración del voto en los cuatro partidos más grandes (74,6 por ciento) de los cuales los diputados del PSC representan el 34 por ciento del total regional. En tanto, en la Sierra hay mayor dispersión, toda vez que los cuatro partidos concentran el 61,4 por ciento. Estos datos señalan el carácter más nacional del PSC, mostrándose como el tercer partido en la sierra a poca diferencia del segundo y contando como el primero en la Costa, esto en un análisis de los últimos 20 años de vida democrática.

### **2. Breve análisis económico con la coyuntura política en el país.**

En los últimos 15 años, comenzando desde el gobierno de Rodrigo Borja hasta el gobierno de Gustavo Noboa, el país se ha encontrado sumergido en pagos de deudas, teniendo problemas internos como los bajos salarios y los altos índices de subocupación y desocupación haciendo que los ingresos de la gran mayoría de los ecuatorianos sean exiguos.

La disminución en la producción para el consumo interno, los bajos ingresos que perciben los trabajadores y el alto costo de la vida sumergen en la pobreza a la población de más bajos recursos, aun cambiando de moneda entendiéndose de Sucres a Dólares de los Estados Unidos de Norteamérica, la economía se ha estabilizado pero los problemas persisten ocasionando que los ecuatorianos emigren. Esto es, que en los últimos gobiernos se los acusa de corrupción, todo esto por malas administraciones, dando síntomas en la moneda local desde el Gobierno de Sixto Durán Ballén y con una inflación que iba en aumento, agudizándose en el gobierno de Abdala Bucaram y finalmente dándole paso a la dolarización con que murió la moneda nacional.

Es que los malos gobiernos han hecho que el Ecuador se encuentre en una posición no favorable frente a los organismos de crédito internacionales, lo que implica que no son flexibles a la hora de otorgar créditos al Ecuador ni que los

inversionistas extranjeros vengan al país, ya que el marco jurídico del Ecuador no es de su total confianza.

### 3. Análisis de Componentes Principales No Lineales o Categóricos.

El Análisis de Componentes Principales No Lineales o Categóricos es parte del análisis multivariante el cual trabaja con variables métricas y variables categóricas o nominales, para este análisis se explicará como actúa el Análisis de Componentes Principales No Lineales o Categóricos, para esto se lo ha dividido en tres partes: la primera trata sobre la codificación de datos categóricos, la segunda trata sobre el concepto de homogeneidad y la tercera nos muestra dos algoritmos que maximizan la homogeneidad de un grupo de variables.

#### 3.1 Codificación de Datos Categóricos

Sea un conjunto de  $n$  objetos o individuos. Una variable  $h_j$  hace corresponder al conjunto de los individuos un conjunto finito de  $k_j$  categorías, este conjunto de categorías se denomina el rango de  $h_j$ . Vamos a asumir que existe un número finito de  $m$  variables  $h_j$  ( $j=1, \dots, m$ ). El producto cartesiano de todas estas categorías se denomina rango multivariante, sus elementos son todas las posibles combinaciones de las  $m$  categorías, y se denominan perfiles. La matriz de datos  $H$  es una matriz  $n \times m$  con elementos  $h_{ij}$  que nos indican la categoría de la variable  $h_j$  para el individuo  $i$ . Estos elementos no necesariamente son números. Un ejemplo de una matriz  $H$  es el siguiente:  $n=10$  (es decir 10 individuos),  $m=3$  (es decir 3 variables),  $k_j=3$  con  $j=1,2,3$  (es decir cada variable tiene tres modalidades). Los elementos de  $H$  son categorías-etiqueta. La primera variable tiene categorías  $A, B, C$ ; la segunda  $P, Q, R$ ; mientras que la tercera  $U, V, W$  (con cero frecuencias para  $W$ ).

$h_1: A, B, C$   $h_2: P, Q, R$   
 $h_3: U, V, W$

H		
B	P	U
	R	V
A	P	U
A	Q	V
B	P	V
C	P	V
A	P	U
A	P	V
C	P	V
A	P	V

##### 3.1.1 VARIABLES INDICATRICES

La matriz  $H$  puede ser codificada utilizando variables indicatrices: Para cada variable  $h_j$  se define una matriz binaria  $G_j$   $n \times k_j$ , de la siguiente forma:

$$g_{(j)ir} = \begin{cases} 1 & \text{si el } i\text{-ésimo individuo está en la } r\text{-ésima categoría de } h_j \\ 0 & \text{caso contrario} \end{cases}$$

$G_j$  se denomina matriz indicatriz de  $h_j$ .

Tales matrices pueden ser reunidas en una matriz particionada  $G=(G_1, \dots, G_j, \dots, G_m)$  de dimensión  $n \times \sum k_j$  también conocida como matriz indicatriz.

### 3.1.2 Cuantificación

Las categorías de las variables pueden ser valores numéricos, como puntos medios de intervalos de alguna variable continua. En este caso la matriz  $H_{n \times m}$  es una matriz de datos clásica y puede ser manejada con las técnicas clásicas del análisis multivariante. En el presente documento no se va a asumir tal cuantificación a priori. Incluso en el caso donde exista tal cuantificación a priori, ésta debe ser ignorada y reemplazada por una categorización nominal. Por ejemplo, si se dispone de la variable edad, esta debe ser dividida en intervalos y a cada uno de éstos asignarle una etiqueta (por ejemplo los puntos medios de cada intervalo):

Supongamos que tenemos una variable “edad” que asigna a los individuos a 15 grupos de edad, cada grupo representado por el punto medio del intervalo en la escala edad. La matriz de datos  $H$  estará formada por una columna con 15 valores. Su correspondiente matriz indicatriz  $G$  en cambio tendrá 15 columnas una por cada grupo de edad. Bien se podría olvidar el origen métrico de estas 15 categorías y pensarlas como 15 categorías nominales.

La cuantificación de categorías sigue ciertas reglas, con la intención de optimizar algún criterio, generalmente este criterio es una función de pérdida. Por el momento no se discutirá tal función, sin embargo se indicará en forma global como la cuantificación de una matriz indicatriz es factible:

La cuantificación de las categorías de la variable  $h_j$  implica que sus  $k_j$  categorías son asignadas como los  $k_j$  valores numéricos de un vector  $y_j$ . Entonces la variable cuantificada  $q_j = G_j y_j$  viene a ser un vector (en  $\mathbb{R}^n$ ) que nos proporciona un resultado numérico para cada individuo con respecto a  $h_j$ .

Definimos  $x$  como el vector promedio de todos los  $q_j$ :

$$x = \frac{1}{m} \sum q_j$$

El vector  $x \in \mathbb{R}^n$  contendrá la cuantificación de los individuos y diremos que para alguna cuantificación directa  $y_j$  de categorías, “ $x$  es el puntaje inducido de los individuos”.

Por otro lado, si  $x$  es alguna cuantificación directa de los individuos, se puede definir una categorización inducida por  $x$  como el promedio de los puntajes de aquellos objetos que asignados en dicha categoría:

$$y_j = D_j^{-1} G_j' x$$

En lo que sigue, se asume que  $D_j$  tiene inversa, lo que significa que no hay categorías con frecuencia cero. Si se este fuese el caso, se debe quitar a esta columna de la matriz indicatriz.

Ambos procedimientos se pueden unir de la siguiente forma:

Sea  $y_j$  una cuantificación directa de las categorías de la  $j$ -ésima variable. Sea  $y$  un vector que esté compuesto por todos los vectores  $y_j$ , es decir tiene  $\sum k_j$  componentes. Los puntajes inducidos de los individuos son:  $Gy/m$ .

Se requiere que una solución para la cuantificación directa de los individuos,  $x$ , sea proporcional a los puntajes inducidos de los individuos y viceversa, que la cuantificación directa de las categorías,  $y_j$ , sea proporcional a la cuantificación inducida de las categorías  $D_j^{-1} G_j' x$ . En la parte 3 veremos dos métodos para obtener soluciones a este problema.

## 3.2 Análisis de Homogeneidad

El término análisis de homogeneidad puede usarse en sentido estricto y amplio. En sentido estricto, denota una técnica para el análisis de datos puramente categóricos con determinada función de pérdida. En sentido amplio, se refiere a una clase de criterio para analizar datos multivariantes en general, compartiendo las características que ayudan a optimizar la homogeneidad de las variables bajo varias formas de manipulación y simplificación.. Esta clase de criterio será utilizado en las próximas secciones.

### 3.2.1 Homogeneidad de Variables

Históricamente, la idea de homogeneidad es cercana a la idea de que diferentes variables pueden medir la “misma cosa”. Si lo último fuese cierto, la matriz de datos (asumiendo que las variables no son otra cosa que desviaciones de su media, es decir que son estandarizadas) podrían dar valores idénticos en cada fila o, si dibujamos las observaciones como perfiles, cada uno de ellos sería una recta. Si la idea de “medir la misma cosa” fuese no fuese muy exacta (las variables miden la “misma cosa” pero con un error aleatorio), las filas de la matriz de datos pueden tener elementos que varían un tanto. Un gráfico de los perfiles sería un línea quebrada.

#### 3.2.1.1 Maximizando la homogeneidad por combinaciones lineales de pesos.

La correlación promedio de las variables nos brinda un estimado de cuan bien ellas pueden ser reducidas a un vector de puntajes, si las mantenemos en su forma original (en este caso en forma estandarizada). Supongamos que es permitido re-escalar a las variables antes de promediarlas, i.e asignar pesos a  $h_j$ , en un intento de incrementar la homogeneidad.

Sea  $x$  un vector de puntajes arbitrario (en  $n$ ) y con media cero. Sea  $a$  un vector de  $m$  pesos. Re-escalar las columnas de  $H$  es equivalente a

reemplazar  $h_j$  por  $a_j h_j$ . El problema que tenemos es el de elegir  $x$  y  $a$  de tal forma que se maximice la homogeneidad. Más explícitamente, minimizar la pérdida de homogeneidad. Por función de pérdida consideramos a la función:

$$\sigma(x, a) \equiv \frac{1}{m} \sum_j SSQ(x - a_j h_j) \quad (2.1)$$

Evidentemente, esta función de pérdida tiene un mínimo absoluto en  $x=0$  y  $a=0$ . Para excluir esta solución trivial es necesario normalizar  $x$  así que  $x^t x = c$  donde  $c$  es una constante dada distinta de 0 (generalmente igual a 1).

El objetivo de elegir puntajes y pesos así como de maximizar la homogeneidad o de minimizar la función de pérdida es una de las posibles definiciones que se puede utilizar para describir a la (primera) componente principal de  $H$ . Esto involucra a combinaciones lineales de pesos, puesto que  $a_j h_j$  pueden ser vistas como una transformación lineal de  $h_j$ . Realmente calcular  $x$  y  $a$  es más complicado que hacer un simple promedio. En la próxima sección veremos que una sucesión de promedios ponderados es suficiente para aproximar la solución tanto como se desee.

### 3.3 Algoritmo de Mínimos Cuadrados Alternantes.

En esta sección se discutirá algoritmos para encontrar puntajes óptimos  $x^*$  y pesos  $a^*$  para el problema de la combinación lineal de pesos. Estos algoritmos se basan en los mínimos cuadrados alternantes. Esto significa que los algoritmos proceden en pasos alternativos, donde en un paso la función de pérdida es minimizada con respecto a  $x$  para un  $a$  fijo y en el siguiente paso la función de pérdida minimizada con respecto a  $a$  para  $x$  fijo. Describiremos dos algoritmos, correspondientes a las dos formas de normalización en conexión con

$$\sigma(x, a) \equiv \frac{1}{m} \sum_j SSQ(x - a_j h_j) \quad (2.1).$$

En el primero  $x$  es normalizado mientras que la escala de  $a$  se deja libre y en el otro  $a$  es normalizado mientras que  $x$  se deja libre. Para mantener una notación sencilla, consideramos que las columnas de la matriz de datos  $H$  son centradas (como variables aleatorias) y están normalizadas (como vectores).

### 3.4 Algoritmo de Puntajes Normalizados.

En este algoritmo los puntajes de los individuos se sujetan a la restricción  $x^t x = 1$ . El algoritmo requiere un vector de pesos inicial y arbitrario  $a^0 \neq 0$ :

1. Actualización de puntajes:  $x^0 \leftarrow \frac{1}{m} H a^0$

2. Normalización:  $x^+ \leftarrow \frac{x^0}{\|x^0\|}$

3. Actualización de pesos:  $a^+ \leftarrow H^t x^+$
4. Test de convergencia: Regresar a (1), hacer que  $a^0 \leftarrow a^+$  mientras los valores de  $x^+$  y  $a^+$  no estén suficientemente estabilizados (de acuerdo a algún criterio de exactitud previamente establecido).

**Descripción del algoritmo:**

1. Corresponde al mínimo condicional no restringido de la función de pérdida (2.1) para un  $a^0$  fijo. Notemos que  $Ha^0/m$  es un vector que contiene los promedios de las filas reescaladas por  $a_j^0$ . Los puntajes actualizados  $x^0$  por consiguiente también minimizan la pérdida relativa  $W/T$  para  $H$  re-escalada con pesos fijos  $a^0$ .
2. Es la proyección de  $x^0$  sobre la hiper-esfera de todos los  $x$  normalizados, lo que transfiere la restricción de minimización a una región factible (la región que contiene todas las soluciones que satisfacen la restricción).
3. Corresponde al mínimo condicional no restringido de la función de pérdida (2.1) para un  $x^+$  fijo. Puesto que  $x^+$  y las columnas de  $H$  son centradas y normalizadas,  $a^+$  es un vector de correlaciones.
4. El algoritmo converge monótonamente, puesto que los pasos 1 y 2 conjuntamente y el paso 3, siempre dan un pequeño valor de la función de pérdida, la cual está acotada inferiormente por 0.

### 3.5 Algoritmo de Pesos Normalizados

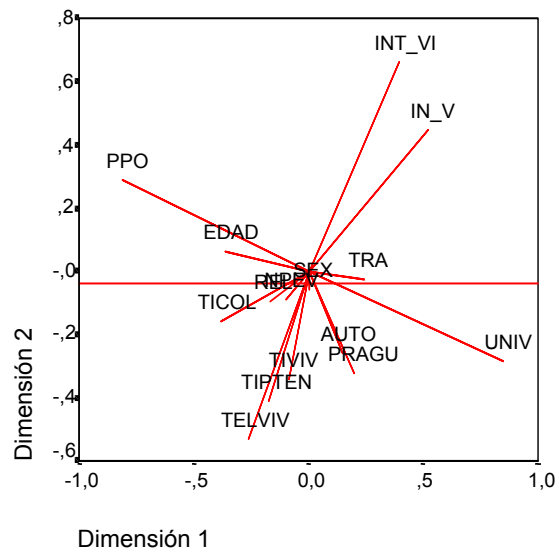
En este algoritmo los pesos satisfacen la restricción  $a^t a = 1$ . Se requiere un valor inicial arbitrario de  $x^0 \neq 0$ :

1. Actualización de pesos:  $a^0 \leftarrow H^t x^0$
2. Normalización:  $a^+ \leftarrow \frac{a^0}{\|a^0\|}$
3. Actualización de puntajes:  $x^+ \leftarrow \frac{1}{m} H a^+$
4. Test de convergencia: Regresar a (1), hacer que  $x^0 \leftarrow x^+$ , mientras los valores de  $x^+$  y  $a^+$  no estén los suficientemente estabilizados (de acuerdo a algún criterio de convergencia previamente definido).

### 3.6 Saturaciones en las componentes

GRÁFICO 1

Saturaciones en las componente



Se observa que las líneas son relativamente largas, indicando que las primeras dos dimensiones explican más de la varianza de todas las variables cuantificadas.

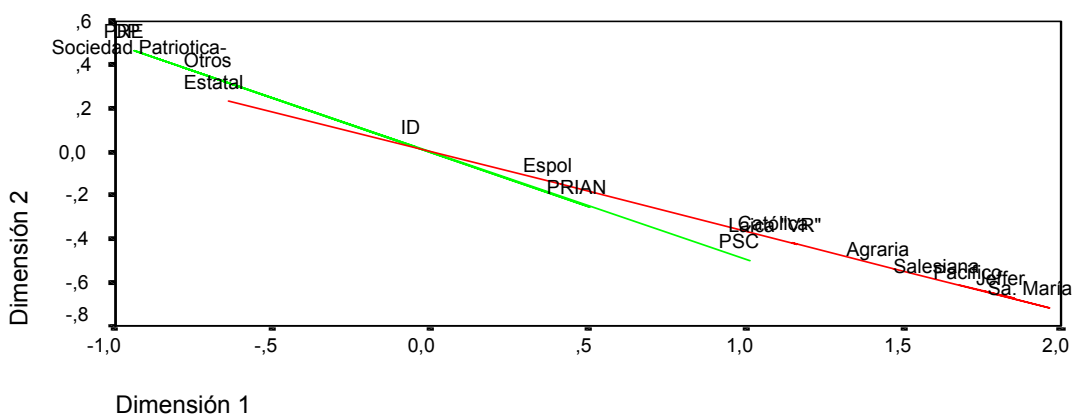
En el gráfico 1 se observan cuatro grupos donde partido político “PPO” es un grupo y universidad “U” es otro grupo los cuales están correlacionados, también Internet e ingresos en la Vivienda es un grupo el cual está correlacionado con el otro grupo donde se encuentran las variables como sexo, edad, religión, teléfono, vehículo, etc, existiendo poca relación con Partido Político y con Universidad.

### 3.7 Saturaciones en las Componentes

En el gráfico observamos el plano en donde se combinan las tres dimensiones, en el cual observamos el comportamiento de las variables.



GRÁFICO 2



**Categorías conjuntas Partido Político \*Universidad**

En el gráfico 2 de categorías conjuntas entre la variable universidad y partido político, se observa como ESPOL, PRIAN y cercanamente la ID forman un grupo, en un extremo inferior derecho se encuentra un grupo donde el PSC se encuentra con las universidades Católica, Laica, Salesiana, Santa María, Jefferson, y la Agraria, mientras que en el extremo superior izquierdo se encuentra un grupo claramente formado por el PRE, DP, Sociedad Patriótica-Pachacutik, y Otros con la Universidad Estatal, se puede notar que la ID se encuentra en medio de dos universidades como la ESPOL y la Universidad Estatal de Guayaquil.

#### 4. CONCLUSIONES

1. Los electores universitarios de la ciudad de Guayaquil, el 75 % de los estudiantes trabajan y el 25 % no trabaja, de donde el 39.9 % proviene de un colegio fiscal, el 31% de particular Religioso, un 26.7 % de un colegio particular laico.
2. El 32 % de los electores universitarios obtuvieron la especialización Informática 29.1%, mientras que el 29.1 % obtuvo la especialización Físico Matemático, siguiendo Químico Biólogo con 17.8 %.
3. Los electores universitarios de la ciudad de Guayaquil, el 80 % de ellos profesa la religión católica, siendo esta la religión mayoritaria entre los estudiantes, seguido de otra con 5.5 %.
4. El 83.9 % de los electores universitarios habita en una casa o villa, de los cuales el 75.48 % tiene un tipo de tenencia propia, el 11.5 % habita en un departamento de los cuales el 4.56 % tiene un tipo de tenencia propia y el 5.77 % tiene un tipo de tenencia arrendada, el 3.6 % habita en cuarto(s) en casa de inquilinato de los cuales el 0.48 % tiene un tipo de tenencia propia y el 2.88 % tiene un tipo de tenencia arrendada, de los electores que habitan en casa o villa el 63 % no trabaja y el 37 % trabaja.
5. El 81 % de los electores universitarios de la ciudad de Guayaquil tiene un tipo de tenencia propia de los cuales el 71,1 % obtiene el agua por tubería dentro de la vivienda, seguido de 14.9 % que tienen un tipo de tenencia arrendada de los cuales el 11.77 % obtiene el agua por tubería dentro de la vivienda.
6. El 83.2 % de los electores universitarios de la ciudad de Guayaquil, dispone de servicio telefónico en el lugar que habita de los cuales el 20.1 % nunca tiene a su disposición un computador con servicio de Internet, mientras que el 25.4 % siempre tiene a su disposición un computador con servicio de Internet, de los cuales el 63.22 % no trabaja.
7. El 40.1 % de los electores universitarios de la ciudad de Guayaquil habita con 4 a 6 familiares, continuándole un 33.7 % de los estudiantes que habita aproximadamente con 2 a 4 miembros en su núcleo familiar, siendo el 32.8 % de los estudiantes investigados tienen ingreso familiar entre el intervalo de USD \$251 a USD \$500, el 21.7 % de los estudiantes investigados tienen ingreso familiar entre el intervalo de USD \$501 a USD \$800.
8. El 72.4 % de los electores universitarios de la ciudad de Guayaquil no posee vehículo de los cuales el 24.5 % tiene una suma de ingresos

familiares USD \$251 a USD \$500, mientras que el 27.4 % posee vehículo.

9. El 61.3 % de los electores universitarios estudian en la universidad de Estatal de Guayaquil, de los cuales el 14.6 % dijo que el partido político de su preferencia es el PRE, seguido de el 9.6 % dijo que el partido político de su preferencia es el PRIAN, el 6.97 % dijo que el partido político de su preferencia es el PSC.
10. El 13.2 % de los electores universitarios estudian en la universidad Laica "Vicente Rocafuerte", el 6.73 % dijo que el partido político de su preferencia es el PSC, seguido de el 5.52 % dijo que el partido político de su preferencia es el PRIAN.
11. El 10.6 % de los electores universitarios estudian en la Escuela Superior Politécnica del Litoral, de los cuales el 36.6 % dijo que el partido político de su preferencia es el PSC, seguido de el 20.45 % dijo que el partido político de su preferencia es el PRIAN y la ID, el 15.9 % dijo que el partido político de su preferencia es *Otros*, y el 6.81 % dijo que el partido político de su preferencia es Sociedad Patriótica- Pachacutik.
12. El 10.3 % de los electores universitarios estudian en la universidad Católica de Santiago de Guayaquil, de los cuales el 67.4 % dijo que el partido político de su preferencia es el PRIAN, seguido del 30.32% dijo que el partido político de su preferencia es PSC, y el 2.3 % dijo que el partido político de su preferencia es Sociedad Patriótica- Pachacutik.
13. Se observó que la variable Partido Político con la variable Edad existe dependencia, es decir que la preferencia de los electores universitarios con respecto al partido político depende de la edad que tenga el elector.
14. Con dos Componentes se explican el 36.805 % de la variación total, haciendo uso de los datos correspondientes a quince variables.
15. En las saturaciones de las componentes se observaron cuatro grupos y existiendo ciertas variables en un vector de medias.
16. Las variables Partido Político y Universidad están correlacionadas, mientras que existe poca relación con Internet en la vivienda y los ingresos mensuales en la vivienda.
17. Las variables como Edad, Género y Trabajo se mantuvieron en una media lo que hacía que no influyan con respecto a las otras variables.
18. En un grupo se encuentran las variables Teléfono, Tipo de vivienda, tenencia de vivienda y número de personas que habitan en la vivienda, existiendo correlación con la variable Internet y la variable ingresos en la vivienda.

## REFERENCIAS BIBLIOGRAFICAS

1. **Baque Bustamante Wilmer**, 2003, "*Estudio de Perfiles socioeconómicos de los electores universitarios de la ciudad de Guayaquil*", Tesis de Grado ESPOL, Guayaquil, Ecuador.
2. **Arias Palacios Hugo**, "*Evolución socioeconómica del Ecuador-Sociedades primitivas y Periodo Colonial*".
3. **Cordes**, "*Guayaquil: Realidades y Desafíos*", S/N Edición, Editorial Ecuador, Ecuador, páginas 122-358.
4. **Douglas C. Montgomery**, "*Diseño y Análisis de Experimentos*", Tercera Edición, Grupo Editorial S.A. de C.V, México, páginas 13-30.
5. **Freund John – Walpole Ronald**, 1990, "*Estadística Matemática con Aplicaciones*", Prentice Hall Hispanoamericana S.A, páginas 449 –453.
6. **Johnson Dallas**, 1998, "*Métodos Multivariados Aplicados al Análisis de Datos*", S/N Edición, International Thonson Editores S. A de C.V., páginas 93-98 y 147-151.
7. **Johnson Richard A. - Wichern Dean**,1998, "*Applied Multivariate Statistical Analysis*", Cuarta Edición, Prentice Hall, U.S.A, páginas 458-460
8. **Macias Washington**,1983, "*Problemas socioeconómicos del Ecuador*", Primera edición, Ecuatextos, Ecuador, páginas 115-220.
9. **Pierre George**, "*Sociología y Geografía*", páginas
10. **Pérez César**, 2000, "*Técnicas de muestreo Estadístico*", S/N Edición, Alfaomega Grupo Editor S.A de C.V, México, páginas 21-25
11. **Universidad Católica**, 1996, "*Cifrando y Descifrando Guayas*", INEC, páginas 51-52.
12. 2001, "*SPSS User's Guide*", SPSS INC. Chicago, USA.
13. 2003, <http://www.puce.edu.ec/Postgrados/MscPoliticlas/Biblioteca/Papel6.htm>, Ecuador.