

## **Análisis Estadístico Multivariante de las Principales Enfermedades Gastroéntéricas en el Cantón Guayaquil**

Cristina Marisela Romero Saltos<sup>1</sup>, John Alex Ramírez Figueroa<sup>2</sup>

<sup>1</sup>Ingeniero en Estadística Informática 2001

<sup>2</sup>Director de Tesis. Matemático, Escuela Superior Politécnica Nacional, 1996, Profesor de ESPOL desde 1995.

### **RESUMEN**

La presente investigación desarrolla el Análisis Multivariado de las principales enfermedades gastroéntéricas que afectan a la ciudad de Guayaquil. Consideramos que las principales enfermedades gastroéntéricas son: Fiebre Tifoidea, Salmonelosis y Enfermedades Diarreicas Agudas.

En el primer capítulo se enfoca la problemática de las enfermedades gastroéntéricas en el Ecuador y en la ciudad de Guayaquil, de la misma manera se presenta la descripción de las principales enfermedades gastroéntéricas y de los agentes biológicos que las provocan.

El segundo capítulo contiene el marco teórico y la descripción de las técnicas univariadas y multivariadas que vamos a utilizar. En el capítulo tres se presenta el análisis univariado. Finalmente el capítulo cuarto contiene el análisis multivariado, la técnica multivariada para la reducción de datos será la de componentes principales.

### **INTRODUCCIÓN**

El desarrollo de este trabajo investigativo está orientada a realizar un análisis estadístico multidimensional, el cual nos permita determinar las relaciones entre las diferentes variables físicas y biológicas de los pacientes que padecen una determinada enfermedad gastroéntérica como es la fiebre tifoidea, salmonelosis y enfermedades diarreicas agudas, así como las asociaciones y grupos que dichas variables. Para determinar dichas asociaciones utilizaremos el método multivariante de reducción de datos denominado Componentes Principales

### **CONTENIDO**

#### **1. La problemática de las enfermedades gastroéntéricas en la ciudad de Guayaquil**

En la ciudad de Guayaquil las condiciones sanitarias y los cuidados en la higiene, son los factores principales que inciden directamente en la problemática de las enfermedades gastroéntéricas. La mayor incidencia de estas enfermedades se reporta generalmente en los meses de invierno, que son los meses en los que se nota la presencia de lluvias, y por la falta de un buen sistema de alcantarillado las aguas lluvias no tienen una libre salida, lo cual provoca el estancamiento de las mismas; esto junto con el hecho de que la mayoría de los agentes patógenos causantes de éstas enfermedades se reproducen con mayor facilidad en aguas estancadas y más aún cuando la temperatura se incrementa.

### 1.1. Enfermedades diarreicas agudas (E.D.A.)

La diarrea es la defecación más o menos líquida y frecuente; se produce por el aumento de secreciones de líquidos en el intestino grueso y una fuerte actividad intestinal. Puede originarse por trastornos nerviosos, por enfriamiento del cuerpo, por infecciones del tubo digestivo, intoxicaciones, abuso de alimentos no imprescindibles, inflamaciones, tumores o enfermedades degenerativas del estomago.

### 1.2. Salmonelosis

La salmonelosis es una enfermedad que puede ser provocada por cualquiera de los aproximadamente 1600 serotipos de salmoenelas. Entre la salmonelosis se distinguen tres tipos:

1. Fiebre Tifoidea, la que la mayoría de los casos es provocada por la Salmonella Typhi.
2. Gastroenteritis aguda, producida generalmente por la Salmoenlla typhimurium.
3. El tipo sépticoemico, que se caracteriza por bacterimia y lesiones fecales, este tipo de salmonelosis es causada más comúnmente por la salmonella choleraesuis.

En este estudio se ha considerado la fiebre tifoidea como un caso diferente de salmonelosis, debido a las marcadas diferencias que presentan.

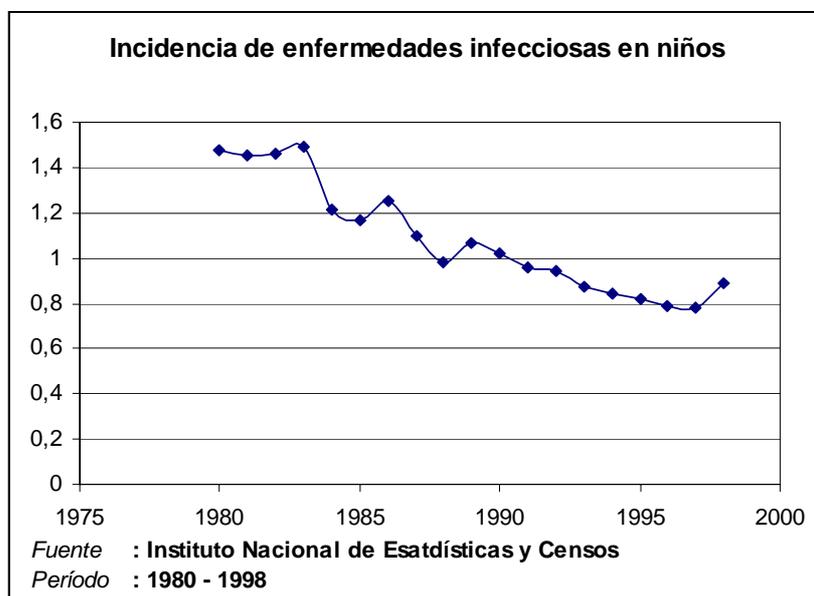
### 1.3. Fiebre Tifoidea

La fiebre tifoidea es una enfermedad que se presenta generalmente en países con bajos niveles de saneamiento ambiental. Es un enfermedad de distribución mundial, se presentan casos sobre todo en muchos países del Lejano y Medio Oriente, el este de Europa, Asia, América Central y América del Sur.

### 1.4. Presencia de las enfermedades gastroéntéricas en el Ecuador

Las enfermedades gastroéntéricas ocupan uno de los primeros lugares entre las diez principales enfermedades que afectan al Ecuador. El porcentaje de incidencia más alto fue en el año 1 983, en este año se reportaron 5 960 casos de niños.

**FIGURA 1**

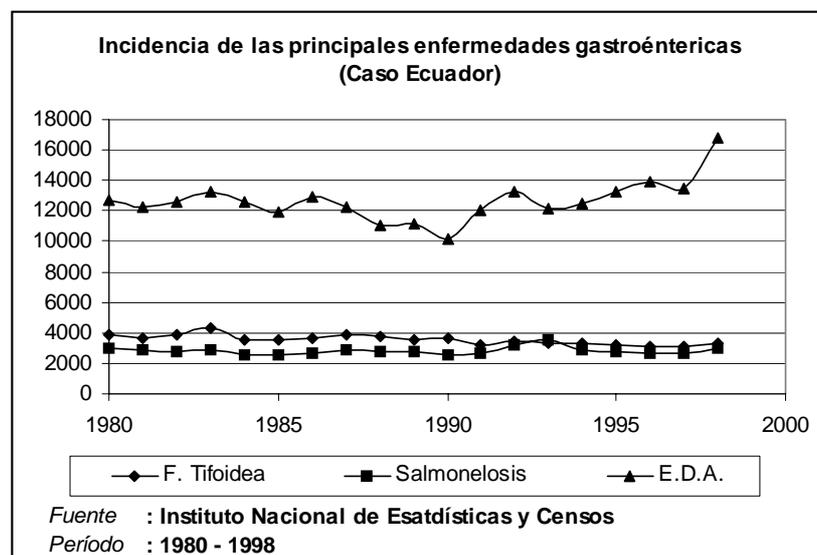


En lo que se refiere al número de casos reportados de niños con fiebre tifoidea, tenemos, que el año en que más casos se reportaron fue en 1 993 el número de casos fue 3 484, mientras que el menor número de casos reportados lo tenemos en el año 1 985 y fueron 2 524 casos.

Con respecto a la presencia de salmonelosis en el Ecuador en el período 1980 – 1998 notamos que el mayor número de casos ocurrió en el año 1983, en este año se reportaron 4 342 casos; en tanto que el menor número de casos fue de 3 249 casos en el año 1991.

En cuanto a las enfermedades diarreicas agudas, el menor número de casos se reportó en el año 1990, en este año se registraron 10 157 casos, mientras que en el año 1998 se registraron 16 734.

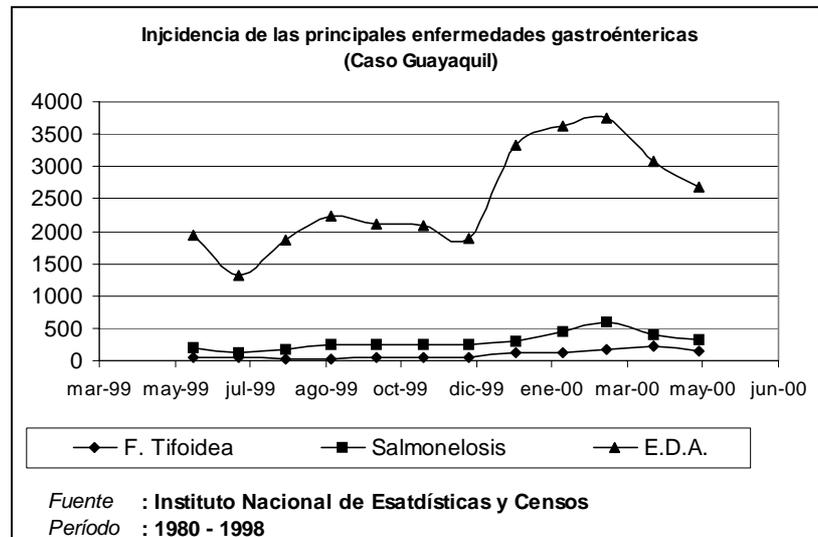
**FIGURA 2**



#### 1.4. Presencia de las principales enfermedades gastréntericas en la ciudad de Guayaquil

En la siguiente figura vemos el comportamiento que han tenido las curvas en las que se han graficado el número de casos que existieron en la ciudad de Guayaquil desde junio de 1999 hasta mayo del 2000, para la fiebre tifoidea, salmonelosis y enfermedades diarreicas agudas.

FIGURA 3



## 2. Determinación de las variables a ser investigadas

Para poder realizar la presente investigación acudimos al hospital del niño }francisco de Ycaza Bustamante de la ciudad de Guayaquil, y tomamos los datos de los pacientes que ingresaron a dicho centro de salud, por presentar un cuadro que se enmarca dentro de las enfermedades gastroéntéricas. Las variables en las que basamos esta investigación son las siguientes:

1. Sexo
2. Edad
3. Peso
4. Estatura
5. Temperatura
6. Pulso
7. Respiración
8. Días de hospitalización
9. Coincidencia del diagnóstico
10. Tipo de parto
11. Establecimiento
12. Edad gestacional
13. Orden de nacimiento
14. Lactancia materna
15. Lactancia compuesta
16. Lactancia entera
17. Tipo de alimentación
18. Desarrollo psicomotor
19. Inmunizaciones
20. Edad de la madre
21. Educación de la madre
22. Ocupación de la madre
23. Edad del padre
24. Educación del padre
25. Ocupación del padre
26. Tipo de vivienda

- 27. Dotación de luz eléctrica
- 28. Abastecimiento de agua
- 29. Eliminación de excretas

### 2.1. Técnica multivariada

Los objetivos relacionados con la explicación de un fenómeno físico o social pueden lograrse recogiendo y analizando los datos. Así al realizar la investigación de algún fenómeno, se debe recoger observaciones de diferentes variables; el método por medio del cual se realiza el análisis de observaciones simultáneas sobre muchas variables es llamado Análisis Multivariado.

Los objetivos de aplicar la técnica multivariada son los siguientes

- Reducción de datos, tanto como sea posible.
- Creación de variables que agrupan objetos o variables similares.
- Investigación de la dependencia entre las variables.

### 2.1. Análisis de Componentes Principales

El análisis de componentes principales está relacionado con la matriz de varianzas y covarianza de un conjunto de variables, a través de algunas combinaciones lineales de esas variables. Las componentes principales son un conjunto de combinaciones lineales de las  $p$  variables aleatorias observadas,  $X_1, X_2, \dots, X_p$ .

### 2.2. Determinación del número óptimo de componentes principales

Para determinar el número óptimo de componentes principales con las que se debe trabajar existen cuatro métodos:

1. El método más utilizado es el implantado por Kaiser (1960). Este criterio consiste en retener solo aquellas componentes cuyos valores sean mayores que 1.
2. El método gráfico denominado Raíces Latentes, el cual fue propuesto por Castell (1966). En este método la magnitud de los valores propios son graficados en el orden en el que fueron obtenidos, generalmente los sucesivos valores propios descienden rápidamente, se recomienda trabajar con las componentes principales correspondientes a los valores propios hasta observar el descenso más pronunciado.
3. El tercer método fue desarrollado por Lawlwy (1940), consiste en realizar una prueba estadística significativa para el número de factores que se deben de retener, sin embargo como todas las pruebas estadísticas, se ve influenciado por el tamaño de la muestra, y un tamaño de muestra grande producirá la retención de un número alto de componentes principales.
4. El último método consiste en retener tantas componentes principales como para contener al menos entre el 80% y el 90% de la varianza total explicada, mediante este método se retienen sólo las variables que son esenciales para las variables especificadas.

## 3. Análisis Estadístico Multivariado

Consideramos que las variables en las que debemos basar la presente investigación son.

- *Variable # 2* : Edad
- *Variable # 3* : Peso

- *Variable # 4* : Estatura
- *Variable # 17* : Tipo de alimentación
- *Variable # 20* : Edad de la madre
- *Variable # 21* : Educación de la madre
- *Variable # 23* : Edad del padre
- *Variable # 24* : Educación del padre
- *Variable # 26* : Tipo de vivienda
- *Variable # 28* : Abastecimiento de agua
- *Variable # 29* : Eliminación de excretas

### 3.1. Análisis estadístico multivariado de las variables observadas para la fiebre tifoidea

La siguiente tabla contiene la varianza que le corresponde a las componentes principales, así mismo contiene el total del porcentaje de explicación. Estos valores se obtuvieron a partir de la matriz de varianza y covarianza de los datos reales.

**TABLA I**  
**PORCENTAJE DE EXPLICACIÓN DE LAS COMPONENTES**  
**(Datos Reales)**

Componente	Varianza	% de explicación	Total % de explicación
1	776,68	86,89%	86,89%
2	63,53	7,11%	94,00%
3	24,78	2,77%	96,77%
4	18,73	2,10%	98,86%
5	4,95	0,55%	99,42%
6	1,84	0,21%	99,62%
7	1,13	0,13%	99,75%
8	0,97	0,11%	99,86%
9	0,62	0,07%	99,93%
10	0,34	0,04%	99,97%
11	0,31	0,03%	100,00%

Notamos que la primera componente principal tiene 86.89% del total de la varianza, lo que indica que al utilizar la primera componente se obtiene el 86.89% del total de la información. La segunda componente proporciona 7.11% de la información total. Al considerar la primera y la segunda componente juntas obtenemos el 94% de la información total.

La primera componente principal es:

$$Y_i = 0.138V(2) + 0.323V(3) + 0.925V(4) - 0.001V(17) + \\ + 0.088V(20) - 0.008V(21) + 0.111V(23) + 0.001V(24) - \\ - 0.008V(26) - 0.004V(28) - 0.001V(29)$$

La variable que predomina en esta componente es la variable número 4, es decir, la estatura.

Cuando las unidades de medida de las diferentes variables no presentan similitud, se debe estandarizar los datos observados. A continuación tenemos la varianza de las once componentes, y de la misma manera se presenta el total del porcentaje de explicación.

**TABLA II**  
**PORCENTAJE DE EXPLICACIÓN DE LAS COMPONENTES**  
**(Datos Estandarizados)**

Componente	Varianza	% de explicación	Total % de explicación
1	3,42	31,03%	31,03%
2	2,32	21,05%	52,09%
3	1,25	11,34%	63,43%
4	1,17	10,62%	74,05%
5	0,85	7,71%	81,76%
6	0,68	6,17%	87,93%
7	0,49	4,45%	92,38%
8	0,42	3,81%	96,19%
9	0,26	2,36%	98,55%
10	0,1	0,91%	99,46%
11	0,06	0,54%	100,00%

Al seguir el criterio de los valores propios mayores que 1, decidimos que el número óptimo de componentes que debemos retener es cuatro, con las cuatro primeras componentes principales obtenemos 74.05% del total de la información original.

Debido a las cargas más altas que presenta cada componente, rotulamos a las cuatro componentes principales de la siguiente manera:

1. Condiciones Biológicas del paciente
2. Entorno Familiar
3. Edad de los padres
4. Condiciones nutricionales

### 3.2. Análisis estadístico multivariado de las variables observadas para la salmonelosis

**TABLA III**  
**PORCENTAJE DE EXPLICACIÓN DE LAS COMPONENTES**  
**(Datos Reales)**

Componente	Varianza	% de explicación	Total % de explicación
1	563,53	75,95%	75,95%
2	120,27	16,21%	92,16%
3	29,24	3,94%	96,10%
4	19,07	2,57%	98,67%
5	3	0,40%	99,08%
6	2,55	0,34%	99,42%
7	1,76	0,24%	99,66%
8	1,13	0,15%	99,81%
9	0,72	0,10%	99,91%
10	0,4	0,05%	99,96%
11	0,28	0,04%	100,00%

La tabla anterior contiene la varianza que le corresponde a las componentes principales, así mismo contiene el total del porcentaje de explicación. Estos valores se obtuvieron a partir de la matriz de varianza y covarianza de los datos reales.

Notamos que la primera componente principal tiene 75.95% del total de la varianza, lo que indica que al utilizar la primera componente se obtiene el 75.95% del total de la información. La segunda componente proporciona 16.21% de la información total. Al considerar la primera y la segunda componente juntas obtenemos el 92.16% de la información total

La primera y segunda componentes principales son:

$$Y_1 = 0.142V(2) + 0.360V(3) + 0.909V(4) - \\ -0.078V(20) - 0.003V(21) + 0.134V(23) - 0.010V(24) + \\ +0.006V(26) - 0.002V(28) - 0.012V(29)$$

$$Y_2 = 0.04V(2) + 0.118V(3) - 0.122V(4) + 0.023V(17) + \\ -0.261V(20) + 0.023V(21) + 0.949V(23) + 0.005V(24) + \\ +0.016V(26) - 0.018V(28) - 0.011V(29)$$

Cuando las unidades de medida de las diferentes variables no presentan similitud, se debe estandarizar los datos observados. A continuación tenemos la varianza de las once componentes, y de la misma manera se presenta el total del porcentaje de explicación.

**TABLA IV**  
**PORCENTAJE DE EXPLICACIÓN DE LAS COMPONENTES**  
**(Datos Estandarizados)**

Componente	Varianza	% de explicación	Total % de explicación
1	3,27	29,70%	29,70%
2	2	18,17%	47,87%
3	1,28	11,63%	59,49%
4	1,17	10,63%	70,12%
5	1,12	10,17%	80,29%
6	0,88	7,99%	88,28%
7	0,46	4,18%	92,46%
8	0,38	3,45%	95,91%
9	0,25	2,27%	98,18%
10	0,13	1,18%	99,36%
11	0,07	0,64%	100,00%

Al seguir el criterio de los valores propios mayores que 1, decidimos que el número óptimo de componentes que debemos retener es cuatro, con las cuatro primeras componentes principales obtenemos 74.05% del total de la información original.

Debido a las cargas más altas que presenta cada componente, rotulamos a las cuatro componentes principales de la siguiente manera:

1. Condiciones Biológicas
2. Edad de los padres

3. Infraestructura sanitaria
4. Condiciones nutricionales
5. Educación de los padres

### 3.3. Análisis estadístico multivariado de las variables observadas para las enfermedades diarreicas agudas (E.D.A.)

De la misma manera que en los casos anteriores, debemos estandarizar los datos observados, para lograr que las unidades de medida de las diferentes variables guarden similitud.

**TABLA V**  
**PORCENTAJE DE EXPLICACIÓN DE LAS COMPONENTES**  
**(Datos Estandarizados)**

Componente	Varianza	% de explicación	Total % de explicación
1	3,84	34,91%	34,91%
2	1,72	15,64%	50,55%
3	1,22	11,09%	61,64%
4	1,2	10,91%	72,55%
5	0,92	8,36%	80,91%
6	0,68	6,18%	87,09%
7	0,54	4,91%	92,00%
8	0,4	3,64%	95,64%
9	0,38	3,45%	99,09%
10	0,06	0,55%	99,64%
11	0,04	0,36%	100,00%

Para el caso de las enfermedades diarreicas agudas, es conveniente retener cuatro componentes principales, ya que con ellas obtenemos el 72.55% de la totalidad de la información. De este modo la cuatro componentes que permiten retener el 72.55% de la totalidad de la información quedan rotuladas de la siguiente manera:

1. Condiciones Biológicas
2. Características Socio – Culturales de los padres
3. Condiciones Socioeconómicas
4. Riesgo de contaminación de los alimentos

### CONCLUSIONES

- El valor promedio para la variable edad en la fiebre tifoidea es de 6.67 años, para la salmonelosis es de 5.53 años y para las enfermedades diarreicas agudas 6.67 años.
- En los casos diagnosticados como fiebre tifoidea el peso mínimo fue de 2.5 kgs. Y el máximo 85 kgs. En cuanto a los casos de salmonelosis el mínimo peso fue de 4.3 kgs. Y el máximo 57 kgs. El valor mínimo para la variable peso en el caso de enfermedades diarreicas agudas fue de 2.5 kgs. Y el máximo 55 kgs.
- En lo que respecta a la variable coincidencia del diagnóstico tenemos que el 52% de las veces el diagnóstico al ingreso fue el mismo que el de egreso. En cambio, en el caso de salmonelosis el 62% de los pacientes fueron bien diagnosticados al ingreso. Mientras, que el 92% de los pacientes que padecieron alguna enfermedad gastroenterica fueron bien diagnosticados.

- La mayoría de las viviendas en las que habitan los pacientes que presentaron alguna de las tres enfermedades (F. Tifoidea, salmonelosis, E.D.A.) habitan en una vivienda construida con hormigón o cemento.
- Con lo que respecta a la variable dotación de luz eléctrica se encontró que el 96%, 97% y 96% de las viviendas en las que habitan los pacientes que enfermaron con fiebre tifoidea, salmonelosis o enfermedades diarreicas agudas, respectivamente, cuentan con este servicio básico.
- En el caso de la información recolectada de los pacientes que adolecieron fiebre tifoidea, puede ser explicada por cuatro componentes principales, condiciones biológicas del paciente, entorno familiar, edad de los padres, condiciones nutricionales.
- Para los datos obtenidos de los pacientes que padecieron salmonelosis, tenemos que cinco componentes principales explican el 74.05% de la información total, éstas son, condiciones biológicas, edad de los padres, infraestructura sanitaria, condiciones nutricionales, educación de los padres.
- En el caso de las enfermedades diarreicas agudas se tiene que la información total puede ser explicada por cuatro componentes, que contienen el 72.55% de la totalidad de la información y éstas son, condiciones biológicas, características socio – culturales de los padres, condiciones socioeconómicas, riesgo de contaminación de los alimentos.
- Se tiene que las principales relaciones que existe entre las variables que se observaron para los casos de las tres enfermedades gastroéntéricas, en las que se ha basado esta investigación, son:
  1. La edad, peso y estatura del paciente mantienen una relación directa.
  2. Específicamente en el caso de las variables observadas para la fiebre tifoidea, se tiene que el tipo de alimentación no presenta una relación significativa con las demás variables.
  3. Específicamente, para el caso de las enfermedades diarreicas agudas, se tiene que la forma de abastecimiento de agua y la forma de eliminar las excretas de la vivienda en la que habita el paciente se relaciona directamente con el tipo de alimentación que recibe el paciente.

## REFERENCIAS

### a) Libro

1. T. W. Drouet, Manual de enfermedades infecciosas y parasitarias (Dirección Técnica – Departamento de Medicina Interna del Hospital de Guayaquil y Universidad Católica de Santiago de Guayaquil)

### b) Libro con referencia

2. R. Jonson, D. Wichern, Applied Multivariate Statistical Analysis (4ta. Edición, New York, Prentice Hall, 1998)

### b) Libro con referencia

3. B. Kliegman, Tratado de Pediatría (15ma. Edición, México, Mc Grawn, 1997)

### b) Libro con referencia

4. S. Pinkus, El Libro de la Mujer (España, Everest, 1993)

### j) Tesis

5. C. Romero, “Análisis Estadístico Multivariado de las Principales Enfermedades Gastroéntéricas en la ciudad Guayaquil” (Tesis, Facultad de Ingeniería en Estadística Informática, Escuela Superior Politécnica del Litoral, 2001)