

“FACTORES PREDICTORES DE SOBREVIDA EN PACIENTES CON DIFERENTES PATOLOGÍAS CANCEROSAS, MEDIANTE EL MODELO DE REGRESIÓN DE COX. CASO: ESTÓMAGO”

Abarca Andrea ¹, Ramírez Jhon ²

Resumen: La población objetivo ha ser analizada son los pacientes de SOLCA pero solo aquellos que fueron diagnosticados con Cáncer de Estómago en el año de 1999, el presente estudio muestra un análisis estadístico de algunas características de la población.

El conjunto correspondiente a la población investigada lo forman 115 pacientes que recibieron atención en la institución ya mencionada. El género masculino tuvo mayor ocurrencia con el 65% que el femenino con el 35%, las edades están entre los 31 a 88 años, con un promedio de 65 años.

El tipo morfológico del cáncer, el 80% fue de tipo intestinal y el 54% de los pacientes tenían metástasis en otros órganos en el cual el 25% fue en el hígado. La mayor parte de los pacientes se encontraban en la fase o estadio IV con el 62%.

En lo relacionado al Análisis Multivariado debemos retener las dos primeras componentes principales, ya que las dos primeras componentes principales explican el 99,998% de la información total.

En lo relacionado al análisis de Supervivencia para el modelo de Regresión de Cox, se analizó dos modelos para contrastar y verificar la información obtenida.

Palabras Claves: Pacientes, Población, Antecedentes, Componentes principales, Regresión de Cox, supervivencia.

1. INTRODUCCION

Con el fin de conocer algunos factores y características de los pacientes atendidos en la Sociedad de Lucha Contra el cáncer o SOLCA diagnosticados durante el año de 1999 con cáncer de estómago.

El objetivo de este estudio es presentar la realidad a la que se enfrentan dicha institución y que con los resultados de este trabajo se pueda mejorar la calidad de los pacientes y ayudar a prevenir a la ciudadanía acerca de este tipo de enfermedad y su temprana detección.

El análisis a efectuarse es estadístico, la técnica de recolección de datos es censal; es decir, se pretendió investigar a todos los miembros de la población objetivo encontrándose que el conjunto correspondiente a la población investigada lo forman 115 pacientes que recibieron atención en esta institución.

2. ANÁLISIS DESCRIPTIVO

Sexo y Edad del Paciente

El género masculino tiene la ocurrencia más alta siendo del 65% mientras que en el femenino es tan solo del 35% de los pacientes. Mientras que en la Edad, el 32% de los pacientes se encuentran dentro del intervalo [64–76), el 25% tenían más de 76 años, un 21% está comprendido entre los [52-64) y el restante 22% tenían menos de 52 años de edad y su media es de 64.69 años.

¹Abarca Andrea, Ingeniera en Estadística Informática; (e-mail: andrea.abarca@hotmail.com)

²Ramírez Jhon, Matemático, Director de Tesis, Coordinador de la carrera de Ingeniería en Estadística Informática, Profesor de la Escuela Superior Politécnica del Litoral (ESPOL) en Instituto de Ciencias Matemáticas; (e-mail: jramirez@goliat.espol.edu.ec).

Lugar de procedencia y nivel de instrucción del paciente.

De los datos analizados el 62% de los pacientes provienen fuera de la ciudad de Guayaquil mientras que solo el 38% reside en esta ciudad. se tiene que la mayor parte de los pacientes, mientras que el 57%, tan solo posee informo poseer educación primaria, el 23% ningún tipo de educación, el 16% educación secundaria y sólo el 4% educación superior

Antecedentes familiares y HPV

Esta bacteria HPV o Hicto Bacto Pílori, al momento de ser diagnosticado se observó que el 83% de los pacientes no poseían ésta bacteria en su organismo mientras que el 17% si poseían y que el 80% de los diagnosticados no tenían ningún familiar que haya padecido de ésta enfermedad a diferencia del 20% que declaró que si había antecedentes de este tipo en su familia.

Clasificación de la lesión.

Esta variable nos permitirá saber el tipo de lugar de la lesión en el estómago que tenía el paciente al momento de ser diagnosticado, de lo cual observamos que el 4% tiene la lesión en el 1/3 proximal cardias, 21% en el 1/3 medio cuerpo, el 15% en el 1/3 distal antro, 51% fue una lesión solapada, es decir en dos o más lugares y 9% fue clasificada como SAI.

Tipo Morfológico

Esta variable nos permitirá a saber cual es el diagnostico según el tipo de morfología que presenta que

presenta el paciente después de los resultados. El 80% de los pacientes tiene del tipo morfológico Intestinal mientras que el 20% de los pacientes del tipo Difuso; lo que quiere decir que en su histología constan con Adenocarcinoma de Tipo Intestinal y Adenocarcinoma de Tipo Difuso correspondientemente.

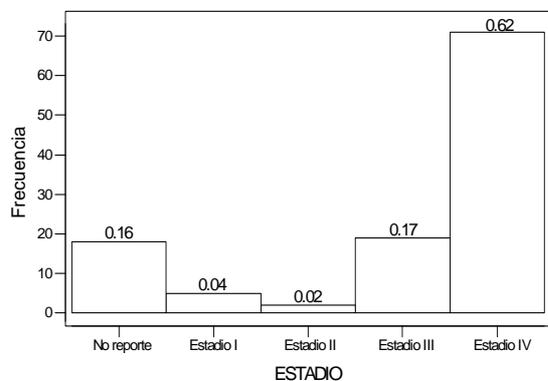
Metástasis y lugar de la metástasis

La Metástasis, da información si el cáncer ha invadido a otros órganos o se encuentra en otra parte del cuerpo que fuera del estómago. Los resultados correspondiente dieron que el 46% de los pacientes el cáncer no se había esparcido por otras partes del cuerpo mientras que, en el 54% si se había esparcido. El lugar donde el cáncer se había esparcido fue de el 25% en el hígado, el 13% posee metástasis carcinomatosis, 3% en los ovarios, huesos, abdomen y/o epiplón cada uno; pulmón y/o pleura, pulmonar y hepática con el 2% cada una y en el 1% se encontró en la Vesícula Biliar.

Estadio

El estadio nos permite identificar en que fase de la enfermedad se encuentran los pacientes. Los resultados que muestra el histograma de frecuencias fueron que en la mayor parte de los pacientes se encontraban en la fase o etapa IV del cáncer con el 62%, en la etapa III el 17%, en la etapa I y en la etapa II el 2%, mientras que el 16% de los pacientes no poseían en el reporte la etapa en que se encontraba su cáncer.

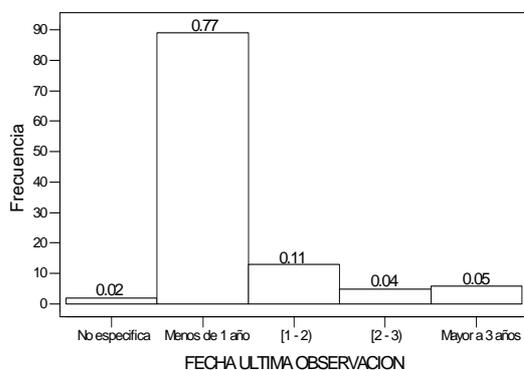
Gráfico I
Histograma de Frecuencia del Estadio



Tiempo de enfermedad.

Aquí deseamos conocer el tiempo transcurrido que asistió al hospital de SOLCA desde que el paciente fue diagnosticado con cáncer gástrico hasta el registro de su última observación. Al analizar el histograma de frecuencias observamos que el 77% de los pacientes asistió a consulta un periodo menor de un año, mientras que el 11% asistió entre un año y dos, el 4% entre dos y tres años y el 5% asistió más de tres años, y un 2% que solo asistió el día que se le diagnosticó.

Gráfico II
Histograma de Frecuencias de Tiempo de Enfermedad



Estado de la Última Observación.

Esta variable nos permitirá saber el estado físico en que se encontraba el paciente en la última cita. El cual al analizar los datos recaudados observamos que la mayoría de los pacientes, el 69%, ya habían fallecidos, el 17% había abandonado el tratamiento o dejó de concurrir al hospital y el 12% se encontraba vivo.

Tratamiento Cronológico

En lo referente a que tipo de tratamiento o tratamientos que recibió el paciente mientras concurrió a SOLCA, observamos que el 63% no recibió ningún tipo de tratamiento, el 25% se le realizó cirugía, el 5% se le realizó cirugía asociada con quimioterapia y el restante se encuentra distribuido entre cirugía, radio y quimioterapia asociadas, quimioterapia, cirugía y radioterapia asociadas y, radio y quimioterapia asociadas.

3. ANALISIS MULTIVARIADO

Análisis de Contingencia

Mediante Tablas de contingencia pudimos observar que las variables Sexo y Estadio son independientes, es decir no tienen influencia la una con la otra.

Entre las variables Tiempo de Enfermedad y Estadio; Nivel de Instrucción y Estado de Última Observación; Tratamiento cronológico y Estadio; Edad y Estadio, pudimos concluir que son variables dependientes es decir si existe dependencia entre estas variables.

ANÁLISIS DE COMPONENTES PRINCIPALES

Se realiza el análisis de componentes principales tomando en cuenta las siguientes variables:

- Edad
- Nivel de Instrucción
- Estadio
- Tiempo de enfermedad

La tabla 1, contiene los valores de la matriz de varianzas y covarianzas, de los datos originales, estos valores representan la relación lineal entre las variables; como se puede observar existe una alta relación inversa entre las variables Edad y Tiempo de Enfermedad también observamos que la variable Tiempo de Enfermedad obtuvo un alto valor de su varianza así como la Edad, por lo tanto estas variables tienen un alto peso en esta matriz.

TABLA 1
MATRIZ DE VARIANZAS Y COVARIANZAS

VARIABLES	Edad	Nivel Instr.	Estadio	Tiempo
Edad	213,18	-0,81	-5,29	-134,29
Nivel Instr.	-0,81	0,57	0,19	-26,26
Estadio	-5,29	0,19	2,22	-35,77
Tiempo	-134,29	-26,26	-35,77	100002,24

Notamos que la primera componente principal tiene 99,767% del total de la varianza, lo que nos indica que utilizar la primera componente obtendremos el 99,767% del total de la información. La segunda componente principal nos proporciona 0,211% de la información total la misma que junto con la primera componente principal nos explica el 99,998% del total de la información.

TABLA 2
PORCENTAJE DE EXPLICACION DE LAS COMPONENTES

λ_i	Varianza	% de explicación	Total % de explicación
λ_1	100896,1	99,787	99,767
λ_2	213,139	0,211	99,998
λ_3	2,101	2,078E-03	99,999
λ_4	0,543	5,368E-04	100,000

La primera componente principal será:

$$Y_1 = -0,029 \text{Edad} - 0,111 \text{Nivel de instrucción} - 0,076 \text{Estadio} + 1 \text{Tiempo de enfermedad}$$

La variable que predomina en esta componente es la variable 15, es decir, el Tiempo de enfermedad.

Análisis de sobrevivencia de las variables observadas por medio del modelo de Regresión de Cox.

Para realizar el modelo de Regresión de Cox plantearemos dos enfoques. En el primero se recodifican las variables Nivel de Instrucción, Tratamiento Cronológico y Estadio en variables dicotómicas y en el segundo enfoque se recodifican las variables Nivel de Instrucción y Estadio usando una escala lickert y Tratamiento Cronológico en una variable dicotómica, mientras que el resto de las variables dentro del modelo permanecen en su codificación original. Al final se hace una comparativa entre los resultados obtenidos mediante estos dos enfoques.

Para el análisis de sobrevivida consideramos junto con el experto que las variables, en las que debemos basar nuestro análisis son:

- Sexo
- Edad
- Nivel de instrucción
- Tiempo de Enfermedad
- Tratamiento cronológico
- Estado de Última Observación (fallecido, vivo o abandono)
- Estadio

Cabe destacar que el Tiempo de enfermedad se lo obtiene de la siguiente forma:

(Fecha de última observación – Fecha de diagnóstico)

Primer Enfoque

Como podemos observar las variables Edad y Tiempo de enfermedad son variables cuantitativas, por lo tanto estas son utilizadas en su forma original, la variable Sexo es una variable de dos categorías, mientras que las variables Nivel de Instrucción, Tratamiento cronológico y Estadio tienen más de dos categorías, por lo tanto estas deben ser recodificadas mediante variables dicotómicas.

**TABLA 3
VARIABLES EN LA ECUACIÓN**

	B	Error estándar	Wald	gl	Sig.	Exp(B)
SEXO	,243	,277	,769	1	,380	1,275
EDAD	,006	,009	,426	1	,514	1,006
N1	-,512	,365	1,966	1	,161	,599
N2	-,270	,438	,381	1	,537	,763
N3	-,459	,755	,369	1	,543	,632
E1	,141	,848	,028	1	,868	1,152
E2	-,172	1,213	,020	1	,887	,842
E3	,582	,647	,809	1	,368	1,789
E4	,866	,407	4,524	1	,033	2,377
T1	-1,414	,486	8,452	1	,004	,243
T2				0		
T3	-,983	1,038	,897	1	,344	,374
T4	,263	,834	,099	1	,753	1,300
T5	-1,712	,646	7,016	1	,008	,180
T6	-13,10	326,369	,002	1	,968	,000
T7	-2,326	1,051	4,898	1	,027	,098

Además la tabla 3 nos indica que los hombres tienen 1.3 veces más probabilidad de contraer éste tipo de cáncer que las mujeres; las personas mayores de 65 años tienen 1.006 veces más que las personas menores de 65 años; de las personas que se encuentran en la Etapa 4 del cáncer tienen 2.3 veces más probabilidad de no superar ésta patología, en la Etapa 3 del cáncer 1.8 veces más mientras que el Etapa 1 es 1.2 veces más.

Como podremos recordar, para determinar si la información proporcionada por la variable X es redundante, se utiliza el p-valor asociado al estadístico de Wald, si una variable es candidata a ser seleccionada en un paso, el criterio de entrada se basa en el p-valor, si este es menor que 0.15 la variable debe ser incluida en el modelo, y si el p-valor es mayor que 0.15 la variable no aporta significativamente en el modelo y por lo tanto debe ser excluida del mismo. Como podemos observar en la Tabla 4-7, las variables Sexo, Edad, Nivel de instrucción (N1,N2,N3), Estadio (E1,E2,E3) y Tratamiento (T2,T3,T4,T6) no deberían incluirse en el modelo debido a que no contribuyen significativamente mientras que la variable Estadio (E4) y Tratamiento (T1,T5,T7) son seleccionadas debido a que si aportan significativamente en el modelo.

Recordemos que, a partir del modelo de regresión de Cox, dado el conjunto de variables independientes $X = \{X_1, \dots, X_p\}$, el límite, cuando Δt tiende a cero, de la probabilidad de que el suceso final ocurra en un

pequeño intervalo $(t, t + \Delta t)$,
 supuesto que no ha ocurrido antes
 del instante t , vendrá dado por:

$$h(t / X) = h_0(t)g(X) = h_0(t)e^z$$

$$\hat{Z} = 0.243\text{Sexo} + 0.006\text{Edad} - 0.512N1 - 0.270N2 - 0.459N3 + 0.141E1 - 0.172E2 + 0.582E3 + 0.866E4 - 1.414T1 - 0.983T3 + 0.263T4 - 1.712T5 - 13.10T6 - 2.326T7$$

siendo \hat{Z} la combinación lineal de las variables

Luego, la estimación de $\hat{g}(X)$ será:

$$\hat{g}(X) = [(1.275)^{\text{Sexo}}(1.006)^{\text{Edad}}] * [(0.599)^{N1}(0.763)^{N2}(0.632)^{N3}] * [(1.152)^{E1}(0.842)^{E2}(1.789)^{E3}(2.377)^{E4}] * [(0.243)^{T1}(0.374)^{T3}(1.3)^{T4}(0.18)^{T5}(0.0)^{T6}(0.098)^{T7}]$$

Esto es para cualquier valor que pueden tomar las variables que se encuentran dentro del modelo. Por ejemplo para un paciente cuyo sexo es masculino, edad es 65 años, su nivel de instrucción primario, con un Estadio 3 y con tratamiento cirugía, tendremos el siguiente resultado:

$$\hat{g}(X) = (1.275)^0(1.006)^{65}(0.599)^1(0.763)^0(0.632)^0(1.152)^0(0.842)^0(1.789)^1(2.377)^0(0.243)^1(0.374)^0(1.3)^0(0.18)^0(0.0)^0(0.098)^0$$

$$\hat{g}(X) = 0,38416 \text{ y}$$

$$h(t/X) = h_0(t)0,38416$$

En el ejemplo dado, el valor obtenido nos dice que el paciente con las características dadas tiene el 38.416% de probabilidad de sobrevida.

TABLA 4
TABLA DE SOBREVIDA

Tiempo	Evento acumulado	Sobrevida	Error Estándar
1	1	0,994	0,018
5	2	0,988	0,036
6	3	0,981	0,054
33	18	0,871	0,399
34	20	0,855	0,379
70	32	0,744	0,619
245	54	0,464	0,998
258	55	0,449	1,007
300	60	0,366	1,024
1159	76	0,016	0,170
1175	77	0,005	0,072
1210	78	0,001	0,023
1349	79	0,000	0,003

Segundo Enfoque

Para realizar el modelo de Regresión de Cox, en este enfoque se recodificó las variable Tratamiento Cronológico en una variable dicotómica y las variables Nivel de Instrucción y Estadio en variables en escala licker, además de que en el proceso se eliminó 18 datos de la variable Estadio correspondiente a la categoría No Reporte para una mejor interpretación. Y como se mencionó en el anterior enfoque las variables Edad y Tiempo de enfermedad son variables cuantitativas, por lo tanto estas son utilizadas en su forma original, la variable Sexo es una variable de dos categorías.

**TABLA 5
VARIABLES EN LA ECUACIÓN**

	B	Error estándar	Wald	gl	Sig.	Exp(B)
SEXO	,448	,288	2,418	1	,120	1,565
EDAD	,007	,009	,519	1	,471	1,007
INSTRUCC	-,050	,211	,056	1	,813	,951
ESTADIO	,169	,237	,512	1	,474	1,185
T1	-1,436	,421	11,661	1	,001	,238
T2				0		
T3	-1,026	1,037	,980	1	,322	,358
T4	,210	,773	,074	1	,786	1,234
T5	-1,633	,580	7,916	1	,005	,195
T6	-13,126	336,134	,002	1	,969	,000
T7	-2,408	1,045	5,304	1	,021	,090

También la tabla 5 nos indica que los hombres tiene 1.6 veces más probabilidad de contraer cáncer de estómago, las personas mayores de 65 años de contraer es 1.007 veces más que las personas menores de 65 años y las personas con Estadio 4 tienen 1.18 veces más probabilidad de no superar ésta enfermedad.

Al igual que el enfoque anterior se utiliza el mismo criterio para determinar si la información proporcionada por la variable X es redundante, Como podemos observar en la Tabla 5, las variables Edad, Nivel de instrucción, Estadio y Tratamiento (T2,T3,T4,T6) no deberían incluirse en el modelo debido a que no contribuyen significativamente mientras que las variables Sexo y Tratamiento (T1,T5,T7) son seleccionadas debido a que si aportan significativamente en el modelo.

Y el modelo vendría dado por:

$$\hat{g}(X) = [(1.565)^{\text{Sexo}} (1.007)^{\text{Edad}} (0.951)^{\text{Nivel de instrucción}} (1.185)^{\text{Estadio}}] * [(0.238)^{T1} (0.358)^{T3} (1.234)^{T4} (0.195)^{T5} (0.00)^{T6} (0.090)^{T7}]$$

Esto es para cualquier valor que pueden tomar las variables que se

encuentran dentro del modelo. Usando el mismo ejemplo anterior, es decir para un paciente cuyo sexo es masculino, edad es 65 años, su nivel de instrucción primario, con un Estadio 3 y con tratamiento cirugía, tendremos el siguiente resultado:

$$h(t|X) = h_0(t) 0,50016$$

Este valor nos indica que el paciente con las características dadas tiene el 50.016% de probabilidad de sobrevivida.

**TABLA 6
TABLA DE SOBREVIDA**

Tiempo	Evento acumulado	Sobrevida	Error Estándar
1	1	0,993	0,024
5	2	0,986	0,049
6	3	0,979	0,073
33	16	0,871	0,414
34	18	0,852	0,468
70	28	0,749	0,743
245	47	0,477	1,205
261	48	0,460	1,217
300	52	0,384	1,245
1159	67	0,027	0,310
1175	68	0,011	0,157
1210	69	0,003	0,056
1349	70	0,000	0,007

Comparación de los modelos

En la tabla de sobrevivida aumento levemente la probabilidad de no fallecer en un tiempo específico.

Como se puede recavar, en ambos modelos se excluyen las variables Edad, Nivel de Instrucción, Estadio y Tratamiento (T2,T3,T4,T6) y se incluyen en el modelo las variable Tratamiento (T1,T5,T7), en lo que se diferencian es en que en el primero excluye la variable Sexo del modelo mientras que en el segundo no,

aparte de que en el primero si incluye la variable dicotómica Estadio (E4).

En el ejemplo realizado con las mismas características de un

paciente se tiene que la función $\hat{g}(X)$ varía de 0.38416 a 0.50016, lo que nos indica que un mismo paciente en el segundo modelo tiene mayor probabilidad de sobrevivir que en el primer modelo.

Por lo tanto la variable Nivel de Instrucción no influye en el modelo mientras que la variable Estadio dicotómica si influye.

4. CONCLUSIONES

1. El género masculino tuvo mayor ocurrencia con el 65% que el femenino con el 35%, en el diagnóstico registrado en el año 1999 y la edad mínima observada fue de 31 años, mientras que la edad máxima fue de 88 años; en el intervalo [64–76) de edad, fue donde se encontraron la mayor cantidad de pacientes con el 32% del total de toda la población analizada con cáncer gástrico.
2. El 83% de los pacientes no poseían en su organismo la bacteria Hicto Bacto Pílori o HPV y el 80% no poseía historia médica en su familia que haya sido detectado con cáncer gástrico.
3. El momento de ser diagnosticado, el 51% de los pacientes tenía una lesión solapada y el 21% en el 1/3 medio cuerpo del estómago. El tipo morfológico del cáncer, el 80% fue de tipo intestinal y el 54% de los pacientes tenían metástasis en otros órganos en el cual el 25% fue en el hígado.
4. La mayor parte de los pacientes se encontraban en la fase o estadio IV con el 62%, en el estadio III el 17%, en el estadio I el 2%; mientras que el 16% no poseían reporte en el estadio en que se encontraba.
5. En la variable Tiempo de Enfermedad, pudimos observar que el 77% de los pacientes que presentan cáncer de estómago asistió a consulta en un periodo menor de un año, el 11% asistió entre un año y dos, el 4% entre dos y tres años y el 5% concurrió más de tres años a consulta.
6. El 63% de los pacientes no recibió ningún tipo de tratamiento cronológico para combatir éste tipo de cáncer, mientras que el 25% recibió cirugía como tratamiento de la cual gastrectomía total fue la que más se utilizó.
7. Mediante Tablas de contingencia pudimos observar que las variables Sexo y Estadio son independientes, es decir no tienen influencia la una con la otra.
9. Al trabajar con la matriz de varianzas y covarianzas de los datos reales debemos retener las dos primeras componentes principales, ya que las dos primeras

componentes principales explican el 99,998% de la información total.

10. El Tiempo de Enfermedad es la variable que mas pesa dentro del modelo y debería ser la primera en considerarse a la hora de realizar un análisis de cáncer gástrico.

11. Para el modelo de Regresión de Cox, se utilizaron todas las variables anteriormente mencionadas, y se utilizó como variable dependiente el Tiempo de sobrevivida, y como variable de estado se utilizó la variable Estado de Última Observación para el evento "fallecido" se obtuvo el siguiente modelo:

$$\hat{Z} = 0.243\text{Sexo} + 0.006\text{Edad} - 0.512N1 - 0.270N2 - 0.459N3 + 0.141E1 - 0.172E2 + 0.582E3 + 0.866E4 - 1.414T1 - 0.983T3 + 0.263T4 - 1.712T5 - 13.10T6 - 2.326T7$$

12. Al comparar ambos modelos, la variable que es significativa en el modelo es el Tipo de Tratamiento (T3, T5, T7).

13. La variable Estadio (E4) es significativa en el primer modelo mientras que la variable Sexo lo es en el segundo modelo, esta discrepancia se debe a que en el primer modelo el número de casos correspondiente al evento Muerte fue de 79 mientras que en el segundo fue de 70.

REFERENCIAS BIBLIOGRAFICAS

1. **JOHNSON, R AND WICHERN, D** (1998). "Applied Multivariate Statistical Analysis", Prentice Hall, Upper Saddle River, New Jersey, USA.

2. **JOHNSON, D.** (1998). "Métodos Multivariados Aplicados al Análisis de Datos", International Thompson Editores, México, México.

3. **VISAUTA, V.** (1997). "Análisis Estadístico con SPSS para Windows. Estadística Básica", McGraw – Hill / Interamericana S.A. Madrid, España.

4. **MENDENHALL, W.** (1994). "Estadística Matemática con Aplicaciones", Grupo Editorial Iberoamérica, México, México.

5. **FREUND J., WALPOLE E.** (1990) "Estadística Matemática con Aplicaciones", Cuarta edición, Prentice Hall / Hispanoamericana S.A. México, México.

6. **CANAVOS, G.** "Probabilidad y Estadística Aplicaciones y Métodos", McGraw-Hill/ Interamericana S.A. México, México.

7. www.press2.nci.nhi.gov/sciencebehind