

"Detección y análisis estadístico de los principales factores que inciden en el cáncer al pulmón"

Por: Juan Javier Laínez Bolaños¹.

Coautor: Mat. John Ramírez².

¹ **Ingeniero en Estadística Informática.**

² **Director de tesis, Profesor del Instituto de Ciencias Matemáticas de la ESPOL.**

Introducción

En el presente trabajo de tesis presentaremos un estudio en el que nosotros vamos a detectar y analizar con herramientas estadísticas factores que inciden en el cáncer al pulmón, para esto nosotros hemos escogido a las personas que fueron internas en el Hospital de SOLCA de guayaquil en el año de 1.997 por este mal.

Primero vamos a dar una breve introducción al respecto al cáncer al pulmón en el Ecuador, luego procedemos a explicar el fundamento teórico las técnicas estadística utilizadas en este trabajo después estas técnicas la aplicaremos en nuestra muestra escogida y las conclusiones mostraran los factores que inciden en este cáncer y las recomendaciones daran la pauta para un mejor estudio

Resumen

En este resumen se describe brevemente cada capitulo en nuestro trabajo de tesis.

Primero tenemos la descripción del cáncer al pulmón en el Ecuador pero primero veremos brevemente todas las incidencias, características, tratamientos para contrarrestar este mal y contagios del cáncer al pulmón.

El fundamento teórico de las técnicas que se aplicaran en el estudio. Básicamente nos dirigimos al Análisis de Componentes Principales(A.C.P.) en la cual se describe toda la base teórica del mismo

La descripción estadística descriptiva y multivariante de las variables que se han escogido en el estudio y presentamos el análisis de los resultado las estadísticas de las variables escogidas.

El Cáncer al Pulmón en el Ecuador

Cáncer al Pulmón

El cáncer es el causante de alrededor del 21 % de todas las muertes, estas son suposiciones que tiene algunos de los diferentes países tanto Europeos como Americanos.

En los países occidentales como el Ecuador el cáncer pulmonar ha tenido un crecimiento significativo en el siglo veinte.

De todas las patologías, el cáncer al pulmón es el que tiene una relación muy fuerte con el tabaquismo, la mayor parte de personas que tiene esta enfermedad son causadas por el consumo del tabaco, esto es aproximadamente el 90% de los enfermos de la misma.

Factores más Importantes para Contraer Cáncer al Pulmón

Algunos de los factores más importantes para contraer esta enfermedad son los siguientes:

- Factores Laborales.
- Contaminación Ambiental.
- Tabaquismo.

Factores Laborales

Este factor es el responsable aproximadamente del 0.5% de las muertes por cáncer. Esto es debido por el aumento de la tecnología y la instauración efectiva de medidas protectoras.

Actividades como las siguientes son las que están implicadas en la obtención de esta patología.

- Trabajadores en minas y canteras.
- Trabajadores en fundición (barros de ladrillos y metales).
- Soldadores.
- Conductores de transportes públicos y camiones.
- Mecánicos de automóviles.
- Electricistas.
- Mecánicos aeronáuticos.
- Pintores.
- Yeseros.
- Trabajadores de la construcción e industrias de: plástico, cuero, vidrio, muebles, goma, material fotográfico.

Carcinógenos conocidos.

- Arsénico.
- Asbesto.
- Bis - clorometil.
- Cromo.
- Níquel.
- Componentes aromáticos policíclicos.
- Radon.
- Vinil - cloride.

Carcinógenos probables.

- Acrilonitrilo.
- Berilio.
- Cadmio.
- Fornaldeida.

Carcinógenos posibles.

- Acetaldeido.
- Fibras sintéticas.
- Silica.

Contaminación Ambiental

Alrededor del 10% del total de los cánceres pulmonares son ocasionados por este factor. Si bien es cierto los estudios realizados son hechos ha conveniencia ya que no han tomado en cuenta algunos aspectos que distorsionan esta información tales como:

- Puestos laborales.
- Hábitos.
- Y otros.

Sin embargo, es razonable suponer que la contaminación industrial, por automóviles, entre otros, debe desempeñar algún papel, aunque éste no haya sido valorado convenientemente aún.

Tabaquismo

Está es una de las mayores causas de las muertes por cáncer en los países occidentales, esto es aproximadamente el 30% de las personas que sufren de esta enfermedad.

Como el consumo habitual del cigarrillo es el síntoma antecedente de casi el 95% del cáncer pulmonar y, además, está presente en otras importantes enfermedades bronquiales, cardíacas, vesicales, entre otros, es importante exponer algo acerca de la dependencia, ligada a lo emocional, que hace que, el "uso social", despreocupado, que se le da al cigarrillo, en un primer momento por curiosidad o porque "lo hace todo el mundo", se convierta en un acto repetitivo, impulsivo, casi imposible de evitar en la mayoría de los fumadores.

El empleo repetitivo de una sustancia puede condicionar a un sujeto a habituarse a ella; más todavía está le provoca un efecto estimulante, haciendo que la situaciones se vuelvan más tolerables. Esto se debe ha que se produce un desplazamiento de la tensión y de la atención a dicha sustancia, en este caso el cigarrillo, lo que permite ocultar los temores y ansiedades y se constituye en sedante parcial para el consumidor.

Enfoque estadístico multivariante para la incidencia del cáncer al pulmón.

Análisis de Componentes Principales. (A.C.P.)

Este método es muy importante, ya que permite la estructuración de un conjunto de datos multivariados obtenidos de una población, cuya distribución de probabilidades es desconocida.

Cabe recalcar que gracias a las características sobresalientes del análisis de componentes principales para efecto de este estudio ayudará a:

- Generar nuevas variables que puedan expresar la información contenida en el conjunto original de datos.
- Reducir la dimensionalidad del problema de este estudio, para realizar un mejor análisis en el futuro.
- Eliminar, cuando sea posible, algunas de las variables originales, siempre y cuando ellas aporten poca información.

Los pasos en el Análisis de los Componentes

Sea M una matriz simétrica, definida y positiva para p filas y p columnas. Esta matriz define un producto escalar sobre el espacio \mathfrak{R}^p considerando en su base canónica. Un primer teorema va a permitirnos considerar indiferentemente el producto escalar o combinación lineal.

Teorema 1.- Sea una v. a. r. C combinación lineal de v. a. r. X_1, \dots, X_p . Entonces, existe un vector u de \mathfrak{R}^p tal como la v. a. r. C sea tal que:

$$\forall \omega \in \Omega \quad C(\omega) = u \cdot X(\omega)$$

$X(\omega)$ representa el vector de \mathfrak{R}^p $(X_1(\omega), \dots, X_j(\omega), \dots, X_p(\omega))$ y el producto escalar siendo definido por la matriz M .

Se planteará por definición: $C = u \cdot X$.

Entonces C es combinación lineal de v. a. r. $X_1, \dots, X_j, \dots, X_p$.

Investigar una v. a. r. bajo forma de combinación lineal o de producto escalar después es equivalente.

Llamamos primera componente principal del vector aleatorio $(X_1, \dots, X_j, \dots, X_p)$ la v. a. r. C_1 solución del problema siguiente:

$$\begin{cases} \max & E(C_1^2) \\ C_1 & = u_1 \cdot X \\ u_1 \cdot u_1 & = \|u_1\|_M^2 = 1 \end{cases}$$

Llamamos segunda componente principal del vector aleatorio X la v. a. r. C_2 solución del problema siguiente:

$$\begin{cases} \max & E(C_2^2) \\ C_2 &= u_2 \cdot X \\ \|u_2\|_M &= 1 \\ E(C_1 C_2) &= 0 \end{cases}$$

Vemos aquí que la segunda componente principal no está correlacionada a la primera.

Llamamos l-ésima componente principal v. a. r. solución del problema siguiente:

$$\begin{cases} \max & E(C_l^2) \\ C_l &= u_l \cdot X \\ \|u_l\|_M &= 1 \\ E(C_l C_{l'}) &= 0 \end{cases}$$

Sometemos luego a la l-ésima componente principal a no estar correlacionada a todos los precedentes.

Sujeto a las restricciones anteriores, se puede demostrar que:

$$C_l = u_l \cdot X$$

Donde U_i es el i-ésimo vector propio de la matriz de Datos

Interpretación de los valores propios.

En el análisis de componentes principales es necesario calcular e interpretar los valores propios generados, ya que de esta forma se podrá decidir cuántos valores propios serán considerados si se desea reducir la dimensión original de las p variables a m variables (siendo $m < p$).

$$\sum_{j=1}^p s_{jj} = \sum_{k=1}^p \lambda_k \quad j, k = 1, \dots, p,$$

Es importante recalcar que cada componente principal explica una proporción de variabilidad total y esa proporción se puede calcular mediante el cociente entre el valor propio y la traza de S. Este cociente se denomina proporción de la variabilidad total explicada por el k-ésimo componente y se lo calcula así.

$$\frac{\lambda_k}{trS} = \text{variación explicada.}$$

Los valores propios se los ordenará en forma creciente, por lo que es posible seleccionar los primeros m valores propios (siendo $m < p$) y la eficiencia del ajuste de los datos originales por los nuevos m componentes principales estará dada por la proporción de la variación total explicada por la suma de los m primeros valores propios.

$$\frac{\sum_{k=1}^m \lambda_k}{trS} 100 = \text{porcentaje de la variación total}$$

Tanto las dos ultimas ecuación, pueden expresarse en forma de proporción o de porcentaje; en el primer caso, la variación total del conjunto original de datos será la unidad y en el segundo, será 100.

Es posible expresar la variación explicada por cada componente o por los primeros m componentes en función de la suma total de los valores propios de la siguiente forma:

$$\frac{\lambda_k}{\sum_{k=1}^p \lambda_k} 100 = \text{porcentaje de variación explicada por el k-ésimo componente}$$

$$\frac{\sum_{k=1}^m \lambda_k}{\sum_{k=1}^p \lambda_k} 100 = \text{porcentaje de variación explicada por los m primeros componentes}$$

Cuando se consideran todos los componentes, es decir si $m = p$, la proporción de la variación explicada es 1 y el porcentaje es 100%.

Creación en un diseño factorial para la matriz de Datos sobre casos de Cáncer al Pulmón obtenidos en el Hospital SOLCA de Guayaquil en el transcurso del año 1.997

Las variables que hemos recopilado en las diferentes historias clínicas de todas las personas que fueron internadas por cáncer al pulmón en el transcurso del año 1.997, esto quiere decir, todas las personas que fueron internadas por cáncer al pulmón desde el 1 de Enero de 1.997 hasta el 31 de Diciembre de 1.997.

Las variables son las siguientes:

Edad

Sexo

Raza

Estado Civil

Lugar de Procedencia

Ocupación

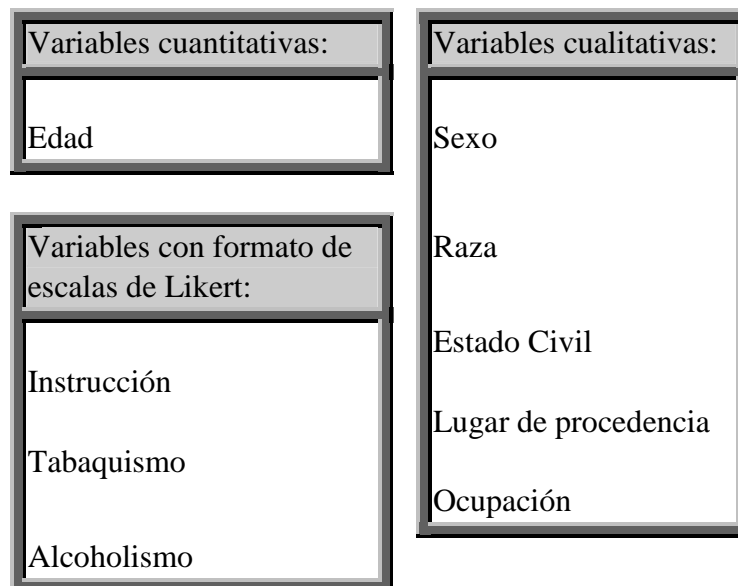
Instrucción

Tabaquismo

Alcoholismo

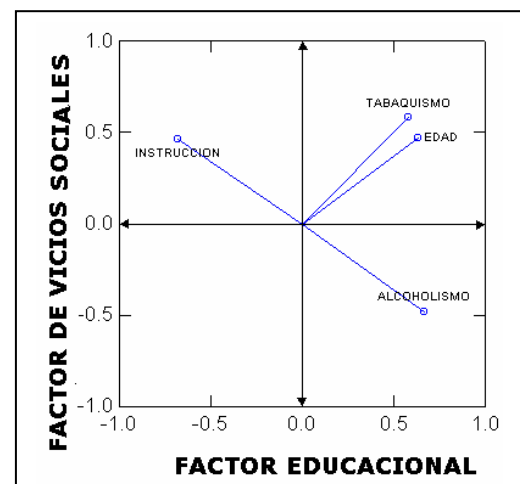
Estadística Multivariante de las variables seleccionadas

La Estadística Multivariante se la llevará acabo con la técnica de los Análisis de los Componentes Principales (A.C.P.) ya explica da en el capitulo dos. Las variables que se utilizarán son la variables cuantitativas y la que está en formato de escalas de Likert.



Usando el A.C.P. obtenemos el siguiente resultado:

Podemos darnos cuenta que la instrucción esta no correlacionada con los vicios y la edad esta correlacionada con el tabaquismo.



Conclusiones y Recomendaciones

Como pudimos ver a lo largo de nuestro estudio hemos empleado técnicas estadísticas univariadas y multivariadas que nos han servido como herramienta para obtener información valiosa sobre los factores que inciden en que una persona obtenga cáncer al pulmón.

En el análisis univariado tuvimos los siguientes resultados:

- La edad promedio que las personas enfermas de cáncer al pulmón que fueron internadas en el transcurso de 1.997 en el Hospital SOLCA de Guayaquil es de 62.58 años.
- El número de las personas internas en el Hospital SOLCA de Guayaquil en su mayoría fueron de sexo masculino con un 69.2% de todos los internos en el intervalo de tiempo ya mencionado, por otro lado el porcentaje de personas de sexo femenino es de 30.8% de todos los internos.
- La raza de los pacientes en dicha época predominaba la raza mestiza tanto es así que el porcentaje de mestizos es de 95.3% y el 4.7% de raza blanca.
- El Estado Civil de las pacientes en intervalo de tiempo ya mencionado en su mayoría eran casado(56% de los pacientes), unión libre(21% de los pacientes), estos son los estados civiles que predominan en esta muestra.
- Las actividades Laborales de los personas cuando fueron internadas eran nulas en un porcentaje significativo(46%), es decir, que una parte significativa de personas no tienen trabajo o han renunciado a su trabajo por asuntos de salud o edad.
- Los internos en su mayoría han recibido instrucción académica(69.2%), el grupo de las personas que han llegado hasta la instrucción secundaria conforman un porcentaje significativo (54% de las personas que han recibido alguna instrucción académica).
- El 53.8% de los internos consumen tabaco, se tienen los siguientes porcentajes de las personas que consumen tabaco:
 - 17.8% consume menos de 15 cigarrillos diarios.
 - 38% consume entre 15 y 25 cigarrillos diarios.
 - 22.1% consumen entre 25 y 30 cigarrillos diarios.
 - 22.1% consume más de 30 cigarrillos diarios.
- Con respecto al consumo de alcohol no tiene mayor porcentaje de consumo(23%).

Estos son los principales resultados univariantes.

Con el análisis multivariado tenemos los siguientes resultados.

- Pudimos darnos cuenta que el promedio de edad de las personas de sexo femenino que fueron internadas es menor(57.9 años) que las personas de sexo masculino(64.6 años).
- También pudimos establecer que el número de personas de sexo masculino que consume tabaco es mayor que el número de personas de sexo femenino, tomando en cuenta que se trata del grupo de personas que tratamos es el ya establecido.
- El consumo de tabaco no esta muy relacionada con el nivel de instrucción, pero no es así con el alcohol que si están inversamente correlacionados de tal forma que se puede decir que a más nivel de instrucción académica menos es el consumo de alcohol.
- En nuestro estudio se había establecido una variable HTA con escala de Likert, esta variable establecía el nivel de hipertensión arterial (HTA)y se notaba my significativa en el estudio, pero al hacer las investigaciones esta variable no es un factor que incide en cáncer al pulmón, si no por el contrario la HTA es causada por el cáncer al pulmón, entonces procedemos a quitarla de nuestro estudio por que no encajaba con nuestros fines.
- Por ultimo al hacer el Análisis de Componentes Principales detectamos los dos principales factores de mi estudio que son los factor educacional y el factor de vicios sociales, y una de las variables que más pesa en el factor de vicios sociales es el tabaquismo.

Y así podemos dar por terminadas las conclusiones más importantes de nuestro estudio.

Como recomendaciones presentamos los siguientes puntos:

- Primero, para un estudio mucho más significativo tenemos que tomar en cuenta factores externos tales como factores laborales o factores climáticos, estos factores no fueron tomados dado que los datos los tomamos de las historias clínicas y no están diseñadas para este tipo de información. Existe una variable que identifica la ocupación de las personas que fueron a internarse por cáncer al pulmón, pero esta variable no tiene la ocupación histórica, es decir, que si la persona ha contraído este mal en su anteriores trabajos no existe esa información en las historias clínicas.
- Segundo, modificar la estructura de las historias clínicas para saber una historia tanto laboral, climática, familiares y personales para tener un punto de vista distinto y así detectar con más certeza los factores que inciden en el cáncer al pulmón. Se debe establecer una escala más amplia para la variable instrucción para saber con certeza la de nivel de instrucción, por ejemplo alguien que solo haya cumplido con el ciclo básico ocupa un nivel de instrucción secundario y esto es falso. Estas extensiones se debe hacer para toda variable de este tipo.

Juan Javier Laínez B.
Graduante

Mat. John Ramírez Figueroa
Director de Tesis

Bibliografía

El Cáncer Broncopulmonar: Generalidades Y Su Relación Con El Tabaco. Universidad De Alcalá De Henares. Escuela Universitaria De Enfermería.
<http://www.interlink.es/enferm/TRABAJO.htmv>

“Oncología“, Volumen 7 (Abril - Julio), Sociedad de Lucha contra el Cáncer (SOLCA)

“Oncología“, Volumen 6 (Septiembre), Sociedad de Lucha contra el Cáncer (SOLCA)

Organización Panamericana de la Salud/Organización Mundial de la Salud (OPS/OMS),
Indicadores de Mortalidad y Morbilidad, <http://www4.salud.org.ec/mor.htm>

Hair Anderson Thathan.Black, Multivariate Data Analisis, Prentice Hall, Unted States of América