

Selección de Modelos en Regresión Logística Binaria Bajo el Enfoqué Clásico: Una aplicación a los problemas de Sobrepeso y Obesidad

MORENO-SANDOVAL, Ma. Isabel*†, DÍAZ-GONZÁLEZ, Lucio``, COVARRUBIAS-MELGAR, Dante``, SISTACHS-VEGA, Vivian del Rosario````

``Unidad Académica de Matemáticas.

``Universidad Autónoma de Guerrero.

````Universidad Autónoma de Guerrero.

````Universidad de la Habana. Ave. Lázaro Cárdenas sin número, Col. La Haciendita, C.U, Chilpancingo, Guerrero. México.

Recibido Junio 4, 2014; Aceptado Octubre 13, 2014

Resumen

Analizando la evolución de este fenómeno en nuestro país se puede constatar que históricamente en México, el principal problema de salud relacionado con la nutrición a principios y mediados del siglo pasado, era la desnutrición, así lo confirman los primeros antecedentes publicados en México que hacen clara referencia a los padecimientos del hambre como problema médico epidemiológico. Para hacer referencia a esto, es el artículo publicado por Patrón Correa en Yucatán (1908) denominado "¿Qué es la culebrilla?". Por otra parte Federico Gómez en (1946) en su artículo "Desnutrición" reconoce que toda la diversidad de los Síndromes referidos anteriormente eran grados de un mismo padecimiento de etiología variada y que actualmente se le conoce como Desnutrición.

Modelos, regresión logística binaria, enfoque clásico.

Abstract

Analyzing the evolution of this phenomenon in our country it can be seen historically in Mexico, the main health problems related to nutrition in early and mid-twentieth century, was malnutrition, confirm the early history published in Mexico that make clear reference to the sufferings of hunger and epidemiological medical problem. To reference this is the article published by Pattern Correa in Yucatan (1908) called "What is shingles?". Moreover Federico Gomez (1946) in his article "Malnutrition" recognizes that the diversity of syndromes referred to above were degrees of the same disease of varied etiology and he is now known as malnutrition.

Models, binary logistic regression, focused classic.

Citación: MORENO-SANDOVAL, Ma. Isabel, DÍAZ-GONZÁLEZ, Lucio, COVARRUBIAS-MELGAR, Dante, SISTACHS-VEGA, Vivian del Rosario. Selección de Modelos en Regresión Logística Binaria Bajo el Enfoqué Clásico: Una aplicación a los problemas de Sobrepeso y Obesidad. Foro de Estudios sobre Guerrero. Mayo 2014 Abril 2015, 1-2:608-612

* Correspondencia al Autor (Correo Electrónico: isabe.mor.san@gmail.com)

† Investigador contribuyendo como primer autor.

Nuevos estudios han documentado que en México, la desnutrición continua siendo un problema de salud pública ya que se encontraba entre las primeras cinco causas de mortalidad infantil, asociado a esto, otros trabajos revelan que la obesidad en México, va en franco aumento.

La Organización Mundial de la Salud (OMS) ha calificado el padecimiento como “la epidemia del siglo XXI”, algunos estudios han registrado el aumento acelerado de la obesidad en la población de todas las edades, tal incremento no es exclusivo de los países económicamente desarrollados, sino que afecta a otras naciones con menor desarrollo, como lo es el caso de México y del resto de países de América Latina.

De acuerdo a estimaciones de la OMS, México ocupa el segundo lugar mundial en obesidad en adultos, después de Estados Unidos. Y por lo que se refiere a la obesidad infantil tiene el primer lugar (INSP, 2009). Es más probable que niños con sobrepeso se conviertan en adultos con obesidad.

En la actualidad, la presencia de esta enfermedad de sobrepeso y la obesidad se da en casi todos los países, este padecimiento se han desarrollado con las condiciones específicas de cada sociedad. Como por ejemplo; “en los países de bajos ingresos la obesidad es más común en mujeres de mediana edad, las personas de mayor nivel socioeconómico y las que viven en comunidades urbanas. Además, en muchos países en desarrollo que se encuentran en fase de transición económica, a menudo conviven en la misma población (o incluso en la misma familia) niveles crecientes de obesidad con desnutrición crónica”.

La hipertensión arterial es una enfermedad que puede evolucionar sin manifestar algún síntoma o ser estos leves, por lo que se le conoce también como “el asesino silencioso”.

Lo que revela que se puede tener hipertensión arterial y sentir se bien. Por lo que, se estima que alrededor del 61% de los individuos afectados desconocen estar enfermos (ENSA 2000).

Objetivos

- Realizar un estudio del problema de la selección de modelos.
- Analizar el uso de los diferentes criterios de selección de modelos bajo el enfoque clásico.
- Realizar un estudio del modelo de regresión logística bajo el enfoque clásico.
- Determinar los factores de riesgo para los problemas, de la hipertensión arterial, sobrepeso y obesidad a través de la selección de modelos en la regresión logística clásica

Metodología

Se estudió una población de individuos que acudían a los diferentes centros de Salud de la Ciudad de Chilpancingo. Los datos fueron recolectados mediante un instrumento de medición, un cuestionario, que se llevó a cabo el 5 de Diciembre del 2011. La muestra del estudio se seleccionó mediante un muestreo no probabilístico, el total de individuos que fueron seleccionados es de 345. Lo cual se quiere determinar los factores de riesgo asociados al sobrepeso y obesidad.

Para la realización del proyecto fue necesario estudiar algunos aspectos fundamentales sobre factor asociado a dichos padecimientos, también el uso y la aplicación del modelo de regresión logística binario, así como los criterios para la selección de modelos aplicados a la regresión binaria.

Dado que los criterios de selección de modelos nos permiten determinar el modelo que se ajuste a los datos.

Los datos fueron analizados en el paquete estadístico R. Las razones de su elección son por la cantidad y calidad de técnicas y funciones implementadas. Así también su manejabilidad, y su versatilidad. Es fácilmente adaptable a una gran variedad de tareas. Además, es libre, accesible y tiene una mayor implantación en la comunidad científica, hasta el punto de considerarse un referente e incluso el estándar para el análisis de datos.

Resultados

Problema 1 (Obesidad).

Se define a la variable Y a los individuos que padecen obesidad:

- 1, Si padece obesidad.
0, No lo padece.

Admitamos también que se consideraron K=9 variables posiblemente influyente en el padecimiento de obesidad.

| Nombre de la variable | Nombre en la base de datos | Descripción de la variable | Tipo de la variable | Valores que toma o unidades de medida |
|-----------------------|----------------------------|----------------------------|---------------------|---|
| Sexo | sexo | Sexo del individuo | Categórica | Mujer
Hombre |
| Edad | edad | Edad del individuo | Continua | Años |
| Estado Civil | estado civil | Estado civil del individuo | Categórica | Casado
Soltero
Divorciado
Viudo
Unión libre
Separado (a)
Hogar |
| Ocupación | ocupación | Ocupación del individuo | Categórica | Comerciante
Chofer
Profesionista
Doctor
Estudiante
Empleado
Plomero
Otro |
| Hipertenso | P3 | Sabe si es hipertenso | Categórica | Si
No |
| Diabético | P9 | Sabe si es diabético | Categórica | Si
No |
| Ejercicio | P18 | Realizar ejercicio | Categórica | Si
No |
| Fumar | P22 | Fumar tabaco | Categórica | Si
No |
| Alcohol | P24 | Consumo de alcohol | Categórica | Si
No |

Tabla 1 Descripción de las variables consideradas para la Obesidad y Sobrepeso.

Se analizaron todas las variables que se consideran influyentes para el padecimiento de la obesidad y se obtuvieron los siguientes resultados.

| Variable | B | Wald | p-valor |
|---------------------------|--------|--------|---------|
| Constante | -1.219 | -1.633 | 0.10246 |
| Edad | 0.023 | 2.371 | 0.01773 |
| Estado civil (Soltero) | -0.823 | -2.003 | 0.04520 |
| Estado civil (Viudo) | -2.406 | -3.014 | 0.00257 |
| Ocupación (Profesionista) | -1.487 | -1.834 | 0.06659 |
| Ocupación (Empleado) | -1.043 | -2.217 | 0.02661 |

Tabla 2 Modelo Completó (1), para la presencia de Obesidad.

| Variable | B | OR | IC 95% OR | p-valor |
|---------------------------|--------|--------|-----------------|---------|
| Constante | -1.010 | 0.3642 | (0.167, 0.7840) | 0.010 |
| Edad | 0.021 | 1.021 | (1.004, 1.040) | 0.017 |
| Estado civil (Soltero) | -0.766 | 0.465 | (0.205, 0.996) | 0.055 |
| Estado civil (Viudo) | -2.314 | 0.099 | (0.015, 0.3842) | 0.003 |
| Ocupación (Profesionista) | -1.375 | 0.252 | (0.038, 1.007) | 0.084 |
| Ocupación (Empleado) | -0.887 | 0.4117 | (0.165, 0.9281) | 0.041 |

Tabla 3 Modelo Reducido (2), solo con las variables que salieron significativas.

Para la variable Estado civil se obtuvo un OR=0.465 en la categoría “Soltero”. Consideremos su inverso obtenemos un OR=2.151 para la categoría “Casado” tomando como referencia “Soltero”, lo que indica que la ventaja a favor de padecer obesidad aumenta 2.151 veces más en individuos que están casados que los que no lo están. Para la categoría “Viudo”, se obtuvo un OR=0.099, considerando su inverso se obtiene un OR=10.10 para la categoría “Casado”, lo que indica que para padecer obesidad aumenta 10.10 veces más en individuos Casados que los que están viudos.

Para la variable Ocupación se obtuvo un OR=0.252 en la categoría “Profesionista”.

Al igual que la variable estado civil, tomemos su inverso y se obtiene un $OR=3.968$ para la categoría “No tener una Profesión”, lo que indica que para padecer obesidad aumenta 3.968 veces más que ser un profesionista. Para la categoría “Empleado” se obtuvo $OR=0.417$, tomemos su inverso y se obtiene un $OR=2.398$, para la categoría “Desempleado”, esto significa que para los individuos sin trabajo aumenta 2.398 veces más probabilidad de padecer obesidad que los que tienen un empleo.

Comparación de los modelos

Modelo 1: $\text{variable_dependiente} \sim \text{edad} + \text{estadocivil} + \text{ocupación} + p3 + p9 + p18 + p21 + p22 + p24$

Modelo 2: $\text{variable_dependiente} \sim \text{edad} + \text{estadocivil} + \text{ocupación}$

| Modelos | Deviance | AIC |
|---------|----------|--------|
| 1 | 405.85 | 447.85 |
| 2 | 412.14 | 438.14 |

Tabla 4 Comparación de los modelos.

Al observar la tabla anterior notamos que, al comparar la Deviance del modelo 1 y el modelo 2, no hay mucha diferencia para elegir cualquiera de los dos modelos, pero con el criterio AIC elegiremos el modelo 2 ya que es el que contiene el menor valor.

Problema 2 (sobrepeso):

Se define a la variable Y a los individuos que padecen sobrepeso:

1, Si padece sobrepeso.
0, No lo padece.

Utilizando las mismas variables de la tabla 1, se tiene el siguiente resultado.

| Variable | B | Wald | p-valor |
|----------------------------|--------|--------|---------|
| Constante | -0.464 | -0.685 | 0.494 |
| Edad | 0.017 | 2.047 | 0.041 |
| Estado Civil (Viudo) | -0.919 | -2.008 | 0.045 |
| Estado Civil (Unión libre) | -0.494 | -1.718 | 0.086 |
| Ocupación (Profesionista) | -0.929 | -1.746 | 0.081 |

Tabla 5 Modelo Completo (1), para la presencia del sobrepeso.

| Variable | B | OR | IC 95% OR | p-valor |
|----------------------------|--------|-------|----------------|---------|
| Constante | -0.224 | 0.799 | (0.414, 1.533) | 0.501 |
| Edad | 0.015 | 1.015 | (0.999, 1.030) | 0.062 |
| Estado civil (Viudo) | -0.971 | 0.376 | (0.153, 0.894) | 0.028 |
| Estado Civil (Unión Libre) | -0.483 | 0.617 | (0.352, 1.072) | 0.088 |

Tabla 6 Modelo Reducido (2), solo con las variables que salieron significativas.

Para la variable estado civil de categoría “Viudo” se obtuvo un $OR=0.376$, tomando su inverso se tiene un $OR=2.66$ para la categoría “Casado”, lo que indica que los individuos casados tienen más probabilidad de padecer sobrepeso que los que no lo están. Para la categoría “Unión libre”, se obtuvo un $OR=0.617$, de igual forma tomando el inverso se tiene que $OR=1.62$ para la categoría “Soltero”, este valor indica que los individuos solteros tienen más probabilidad de padecer sobrepeso.

Modelo 1: $\text{variable_dependiente} \sim \text{edad} + \text{estadocivil} + \text{ocupación} + p3 + p9 + p18 + p21 + p22 + p24$

Modelo 2: $\text{variable_dependiente} \sim \text{edad} + \text{estadocivil} + \text{ocupación}$

| Modelos | Deviance | AIC |
|---------|----------|--------|
| 1 | 634.95 | 682.95 |
| 2 | 642.46 | 668.46 |

Tabla 8 Comparación de los modelos.

Al comparar los dos modelos se observa que el valor de la Deviance del modelo1 y el modelo 2 resultó que no hay mucha diferencia para escoger cuál de los dos modelos utilizar, y lo mismo pasa con el criterio AIC.

Discusión

Los niveles de sobrepeso y obesidad han aumentado progresivamente en las últimas décadas tanto en países desarrollados como en muchos otros en vías de desarrollo. Se estima que más de mil millones de adultos sufren de sobrepeso y por lo menos 300 millones son obesos. En América Latina también se está produciendo un rápido y alarmante incremento en la prevalencia de obesidad, debido principalmente a factores tales como el crecimiento económico, la urbanización progresiva y el consiguiente cambio en los estilos de vida que involucra un mayor consumo de alimentos de alta densidad energética, el abandono de dietas tradicionales y la disminución generalizada de los niveles de actividad física.

La epidemia de sobrepeso y obesidad es preocupante debido a que existe abundante evidencia que indica que el peso excesivo aumenta el riesgo de padecer diversas patologías que alteran significativamente la calidad de vida, entre estas se encuentran las enfermedades cardiovasculares, hipertensión arterial, dislipidemia, insuficiencia cardiaca, resistencia a la insulina, diabetes Mellitus del tipo 2, problemas respiratorios, apnea del sueño, enfermedad vesicular, osteoartritis y diversos tipos de cáncer. Un nivel de presión arterial de 140/90 mmHg o más es considerado hipertensión para la mayoría de adultos.

Conclusión

Para concluir, podemos mencionar que:

El Modelo de Regresión Logística nos sirvió para analizar que variable son importantes en cada uno de los modelos. Ya que los modelos de regresión logística binaria son muy útiles en epidemiología, el ajuste de un modelo puede llevarnos a conclusiones muy importantes como ocurre aquí.

La utilización del software estadístico R es de manera muy eficiente, y además es libre y accesible y mostró ser eficaz en su utilización. De acuerdo a los resultados obtenidos, se concluye que para los dos casos de enfermedades la variable edad es un factor de riesgo para padecer cual quiera de las tres enfermedades.

Los factores de protección fueron:

- Para la obesidad son ser soltero o viudo y tener una profesión o tener un empleo.
- Para el sobrepeso el factor de protección es ser viudo o estar en unión libre.

Referencias

Ávila, Hambre, desnutrición y sociedad: La investigación epidemiológica de la desnutrición en México, México (DF). Ed. Universidad de Guadalajara; 1990. (Colección Fin del Milenio, Serie Medicina Social). 89-175.

Barquera S, Epidemiological and nutritional transition in Mexico: rapid increase of non-communicable chronic diseases and obesity. [versión electrónica]. Public Health Nutrition . 5(1A), 113-122.

Secretaría de Salud (2008). Protocolo Clínico para el diagnóstico y tratamiento de la obesidad. México. México: Secretaría de Salud.

Moreno I. et al. Modelo de Regresión Logística: una aplicación al problema de Hipertensión Arterial, Sobrepeso u Obesidad de personas que acuden a los diferentes Centros de Salud de la Ciudad de Chilpancingo. Matemáticas Estadística, ISB 978-607-487-571-3. BUAP. Año 2013.

Silva L. Excursion a la Regresión Logística en Ciencias de la Salud. Diaz de Santos. 1995.