

# Obesity in Ecuadorian adults. An application of Poisson regression and Logistic regression

Sandra García-Bustos, Ph.D.<sup>1</sup>, Nadia Cárdenas-Escobar M.Sc.<sup>1</sup>, Iris Bustamante Peña<sup>1</sup>, Stefany Uguña Salazar<sup>1</sup>,  
Kevin Peralta Rizzo<sup>1</sup>

<sup>1</sup>Facultad de Ciencias Naturales y Matemáticas, Escuela Superior Politécnica del Litoral,  
Km. 30.5 Vía Perimetral, Guayaquil, Ecuador

*Abstract -- Obesity is a disease which affects around 650 million people worldwide and more than 4.5 million in Ecuador. These figures are alarming because it is recognized as a significant factor in the non-communicable chronic diseases, which appear in all age groups. This work focusses on finding the relations between obesity depicted by Body Mass Index (BMI) and other variables such as gender, province, age group and geographic area for adults between 19 and 59 years old. To achieve this purpose, Logistic and Poisson regression models have been applied, and these results were significant in both models. After contrast pseudo  $R^2$ , the best model has been identified as logistic regression due to its best fit and prediction. As a result, it has been obtained that for an adult male who lives in a rural area of Guayas, the possibility of being obese decreases by 78%. Furthermore, it can be affirmed that if an individual who lives in the province of Guayas moves to the province of Pichincha, and the rest of variables remain constant, then the possibility of being obese declines by 31%.*

**Keywords – Obesity, regression, logistic, Poisson.**

Digital Object Identifier (DOI):  
<http://dx.doi.org/10.18687/LACCEI2019.1.1.401>  
ISBN: 978-0-9993443-6-1 ISSN: 2414-6390

# Análisis de la obesidad en los adultos ecuatorianos. Una aplicación de regresión Poisson y regresión Logística

Sandra García-Bustos, Ph.D.<sup>1</sup>, Nadia Cárdenas-Escobar M.Sc.<sup>1</sup>, Iris Bustamante Peña<sup>1</sup>,  
Stefany Uguña Salazar<sup>1</sup>, Kevin Peralta Rizzo<sup>1</sup>

<sup>1</sup>Facultad de Ciencias Naturales y Matemáticas, Escuela Superior Politécnica del Litoral,  
Km. 30.5 Vía Perimetral, Guayaquil, Ecuador

[slgarcia@espol.edu.ec](mailto:slgarcia@espol.edu.ec), [ncarden@espol.edu.ec](mailto:ncarden@espol.edu.ec), [ijbustam@espol.edu.ec](mailto:ijbustam@espol.edu.ec), [suguna@espol.edu.ec](mailto:suguna@espol.edu.ec), [kperalta@espol.edu.ec](mailto:kperalta@espol.edu.ec)

*Abstract— Obesity is a disease which affects around 650 million people worldwide and more than 4.5 million in Ecuador. These figures are alarming because it is recognized as a significant factor in the non-communicable chronic diseases, which appear in all age groups. This work focusses on finding the relations between obesity depicted by Body Mass Index (BMI) and other variables such as gender, province, age group and geographic area for adults between 19 and 59 years old. To achieve this purpose, Logistic and Poisson regression models have been applied, and these results were significant in both models. After contrast pseudo  $R^2$ , the best model has been identified as logistic regression due to its best fit and prediction. As a result, it has been obtained that for an adult male who lives in a rural area of Guayas, the possibility of being obese decreases by 78%. Furthermore, it can be affirmed that if an individual who lives in the province of Guayas moves to the province of Pichincha, and the rest of variables remain constant, then the possibility of being obese declines by 31%.*

**Keywords—** obesity, regression, logistic, Poisson.

## I. INTRODUCCIÓN

El hábito alimenticio se ha visto afectado por el acelerado cambio que presenta la sociedad en el estilo de vida enfrentándose diariamente a desafíos en el aspecto económico, social y cultural. Es poca o casi nula la atención que se presta a los alimentos que ingerimos, que junto con la reducción de actividad física puede desencadenar en sobrepeso u obesidad. La obesidad se ha convertido en uno de los mayores problemas de salud que afecta a la sociedad, considerándose una enfermedad no transmisible. Además, ésta es catalogada como un factor que produce varios tipos de patologías graves como son la diabetes, hipertensión, enfermedades cardiovasculares, algunos tipos de cáncer y problemas respiratorios [1].

De acuerdo a la Organización Mundial de la Salud [2] esta enfermedad se ha triplicado a nivel mundial en los últimos cuarenta años. En el año 2016, más de 1900 millones de adultos mayores de 18 años tenían sobrepeso, de los cuales, más de 650 millones padecían de obesidad. En general, alrededor del 13% de la población adulta a nivel mundial era obesa, revelando que

el 15% de mujeres y 11% de hombres presentaba este problema. En Ecuador, según el Instituto Nacional de Estadísticas y Censos [3], el sobrepeso y obesidad es un problema que afecta a todos los grupos etarios, considerándose una verdadera epidemia con consecuencias alarmantes en el incremento de las enfermedades crónicas no transmisibles. Resultados obtenidos de la Encuesta Nacional de Salud y Nutrición, ejecutada en el 2012, indican que el 22.2% de habitantes adultos prevalece de obesidad, presentándose en mayor proporción en las mujeres puesto que alcanzan el 27.6%, mientras que los hombres presentan el 16.6%.

Conscientes de la importancia de profundizar en este tema, este estudio se enfoca en la aplicación de las regresiones Logística y Poisson con el objetivo de determinar información relevante acerca del sobrepeso y obesidad en las dos principales provincias del país, como lo son: Pichincha y Guayas. Se ha establecido como objeto de estudio, a los adultos (hombres y mujeres) desde los 19 años hasta los 59 años y se ha considerado como variables relacionadas a grupo de edad, sexo y lugar de residencia.

Se realiza un análisis descriptivo para conocer las posibles variables a incluir en el modelo para posteriormente aplicar ambos modelos y encontrar sus coeficientes. Se determina el ajuste que los modelos de regresión Logística y Poisson tienen con los datos de obesidad para compararlos y elegir el mejor modelo.

## II. MARCO TEÓRICO

### Sobre la obesidad

La Obesidad es una enfermedad compleja y multifactorial [1] que suele iniciarse en la infancia y adolescencia, que se establece por un desequilibrio entre la ingesta y el gasto energético [4].

Esta enfermedad se identifica por la presencia extrema de acumulación de grasa y se manifiesta a través de la abundancia de peso y volumen corporal [5].

Digital Object Identifier (DOI):

<http://dx.doi.org/10.18687/LACCEI2019.1.1.401>

ISBN: 978-0-9993443-6-1 ISSN: 2414-6390

La medida que se utiliza para calcular el sobrepeso y la obesidad es el índice de masa corporal (IMC). Se lo define como un indicador simple de la relación entre el peso y la talla. Se calcula dividiendo el peso de una persona en kilos por el cuadrado de su talla en metros (kg/m<sup>2</sup>). Los adultos con IMC superior a 25 son identificados como personas con sobrepeso y aquellos con más de 30 están encasillados en la obesidad [2].

El objetivo de este trabajo es encontrar las variables que inciden en la obesidad de los ecuatorianos. Se desea encontrar el mejor modelo que pueda predecir el comportamiento de esta enfermedad y para esto se realiza el contraste de regresiones Logística y Poisson.

### Modelo de regresión logística

La regresión logística es un modelo muy aplicado en el estudio de variables categóricas. Es usada en diversas áreas como estadística social, biomedicina, banca, ventas, etc. [6]

Las variables que son medidas pueden ser dicotómicas (dos categorías), nominal (más de dos categorías sin orden inherente mayor o menor que entre las categorías) u ordinal (más de dos categorías con orden inherente mayor o menor entre las categorías). Además, las variables tienen pocas categorías que usualmente son menores que diez. [7].

En regresión logística la variable dependiente puede ser dicotómica (dos categorías), politómica nominal (tres o más categorías con un orden no natural entre las categorías), o politómica ordinal (tres o más categorías con un orden natural entre las categorías). La variable dicotómica dependiente describe la interpretación de una respuesta como “sí” o “no” (1=SI, 0=NO). Ejemplo si alguien usa drogas, si usa computadora, gana la lotería, o si padece de alguna enfermedad. [7].

Para aplicar el modelo logístico es importante entender el concepto de *odds*, cuya representación es la razón de la probabilidad de que un evento ocurra sobre la probabilidad de que el mismo evento no ocurra. La fórmula para encontrarlo se obtiene de p dividido para 1-p, donde p denota la probabilidad del evento de interés.

Si definimos la variable aleatoria binaria X

$$X = \begin{cases} 1, & \text{si el resultado es exitoso} \\ 0, & \text{si el resultado es fracaso} \end{cases} \quad (1)$$

Con la probabilidad de P(X=1)=p y P(X=0)= 1-p [6].

Esto sería una distribución Bernoulli para la cual el E(x) =p, siendo un caso especial de la binomial (1,1) n= 1. La función de masa de probabilidad es [8]:

$$\begin{aligned} f(x, p) &= p^x(1-p)^{1-x} = (1-p) \left[ \frac{p}{(1-p)} \right]^x \\ &= (1-p) \exp \left( x \log \frac{p}{1-p} \right) \end{aligned} \quad (2)$$

El parámetro natural log [p/(1-p)] es el log odds de respuesta 1, es decir el logit de p. Generalmente los modelos que usan la función logit son llamados modelos logit.

El modelo de regresión logística está dado por la siguiente expresión:

$$p(x) = \frac{e^{(\alpha+\beta x)}}{1 + e^{(\alpha+\beta x)}} \quad (3)$$

Como x tiende a infinito, p(x) tiende a cero cuando  $\beta < 0$  y p(x) tiende a uno cuando  $\beta > 0$ .

Encontrando la función de regresión logística para odds, tenemos:

$$\frac{p(x)}{1-p(x)} = e^{(\alpha+\beta x)} \quad (4)$$

Equivalentemente el logaritmo de odds tiene la relación lineal:

$$\log \frac{p(x)}{1-p(x)} = \alpha + \beta x \quad (5)$$

Para el modelo de regresión logístico múltiple donde se tiene k variables predictoras, la expresión es:

$$\log \frac{p(x)}{1-p(x)} = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (6)$$

La fórmula alternativa, especificando directamente p(x) es [8]:

$$p(x) = \frac{e^{(\alpha+\beta_1 x_1+\beta_2 x_2+\dots+\beta_k x_k)}}{1 + e^{(\alpha+\beta_1 x_1+\beta_2 x_2+\dots+\beta_k x_k)}} \quad (7)$$

Dentro del contexto epidemiológico, observamos variables independientes X<sub>1</sub>, X<sub>2</sub> y así sucesivamente hasta X<sub>k</sub> para las cuales se ha determinado un estado referente a una enfermedad, tal que es 1 si “tiene la enfermedad” o 0 si “no tiene la enfermedad”. Se desea describir la probabilidad de que la enfermedad se desarrolle durante un período de estudio. Para modelar esta probabilidad se usa la función logística que la describe en forma matemática de la siguiente manera:

$$P(x) = \frac{1}{(1 + e^{-(\alpha+\sum_{i=1}^k \beta_i X_i)})} \quad (8)$$

Donde X<sub>1</sub>, X<sub>2</sub>, ... X<sub>k</sub> representan las variables independientes o de explicación;  $\alpha$  y  $\beta_i$  representan parámetros desconocidos

del modelo de regresión que necesitaremos estimar con los datos recolectados. [9]

### Modelo de regresión Poisson

El modelo de regresión Poisson se utiliza para caracterizar datos del conteo del número de veces que ha ocurrido cierto evento, estos datos son números naturales, como por ejemplo número de individuos que mueren, número de días de calor intenso o número de individuos que padecen cierta enfermedad. Los datos de tipo conteo se pueden modelar con el uso de Modelo Lineales Generalizados (GLM por sus siglas inglés) con una distribución de errores tipo Poisson. Para esto, se utiliza una función logarítmica cuyo resultado son valores pronosticados positivos, mientras que la distribución de errores de tipo Poisson provee que los datos sean enteros y que la varianza aumente proporcionalmente a la media. [10].

En el clásico modelo lineal, si  $y$  es una variable dependiente y  $X$  es un conjunto de variables independientes tal que  $X_1, X_2, \dots, X_k$  son  $k$  variables diferentes, un modelo de regresión lineal multivariado se define como:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_K X_K + \varepsilon \quad (9)$$

donde  $\beta_0, \beta_1, \dots, \beta_k$  son coeficientes correspondientes a las variables independientes que  $X_1, X_2, \dots, X_k$  y  $\varepsilon$  es el término del error, debido a que con este modelo es casi imposible obtener exactamente todos los valores de  $y$  considerando solo  $\beta$  y los valores de  $X$ . El modelo debe cumplir supuestos importantes que una regresión lineal deben mantener para obtener inferencias válidas de  $\beta$ . Estas son esencialmente cuatro que se incluyen en situaciones cotidiana: linealidad, homocedasticidad, normalidad de los errores, independencia entre  $\varepsilon$  y  $y$ .

Cuando las variables son tipo conteo Poisson, estas condiciones no se cumplen, es por esto que usualmente se realiza un pretratamiento de las variables. Un método posible es transformar en otra variable, que llamaremos  $z$ , entonces el modelo de regresión lineal  $z$  satisface todos los requerimientos. El método de transformación de  $y$  a  $z$  es llamado función vínculo. En la regresión Poisson la función vínculo es una función logarítmica, expresada de la siguiente forma [11]:

$$z = \log(y)$$
$$z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_K X_K + \varepsilon \quad (10)$$

Para estimar los valores de  $\beta$  se usa el método Máxima Verosimilitud y los errores que deben cumplir el supuesto de normalidad son los tipos deviance o Pearson [9].

### III. SUJETOS Y MÉTODOS

Los datos que se utilizan para este análisis son obtenidos de la Encuesta Nacional de Salud y Nutrición, realizada por el Instituto Nacional de Estadísticas y Censos [3]. Esta encuesta recopila información sobre la salud reproductiva, materna e infantil, enfermedades crónicas no transmisibles, estado nutricional, consumo alimenticio, acceso a programas de complementación alimentaria y suplementación para protección de enfermedades, además de datos sobre acceso a los servicios de salud y gasto en salud de la población ecuatoriana.

Se ha tomado los datos de las provincias de Guayas y Pichincha, considerando sólo a los adultos desde los 19 años hasta los 59 años. Se los clasifica por sexo (Hombre y Mujer), luego por grupos de edad: Adulto joven (19 - 35) y adulto (36 - 59) y, por último, de acuerdo a los grupos de IMC (Sobrepeso y Obesidad).

Con esta población definida, se obtiene un total de 2537 personas entre hombres y mujeres para el 1er. grupo de adultos jóvenes y para el 2do. grupo de adultos un total de 2098 personas. Alcanzando un total acumulado de 4635 personas.

Se debe destacar la importancia que tienen las variables dummy, debido a la significación de parámetros y validez del modelo que se comprueba de forma habitual. Es muy importante tener en cuenta que el efecto de una variable debe ser evaluado con todas sus dummies. En otras palabras, las dummies son inseparables en el modelo, no podemos introducir sólo una parte ni eliminar del modelo ninguna variable, cuando la variable está en él. La evaluación del conjunto de dummies puede hacerse a través de los estadísticos y significación del cambio.

Como en regresión logística, las inferencias no requerirán suposición distribucional alguna más allá que la de la propia verosimilitud. Se dirá que las inferencias son asintóticas, es decir, válidas para un  $n$  suficientemente grande.

### IV. RESULTADOS

Para realizar este análisis se utilizaron técnicas estadísticas que nos permitan conocer las variables que pueden impactar en la presencia o ausencia de la obesidad.

Realizando un análisis preliminar o exploratorio de los datos, se determinaron posibles variables a estudiar, como son: edad, sexo y lugar de residencia.

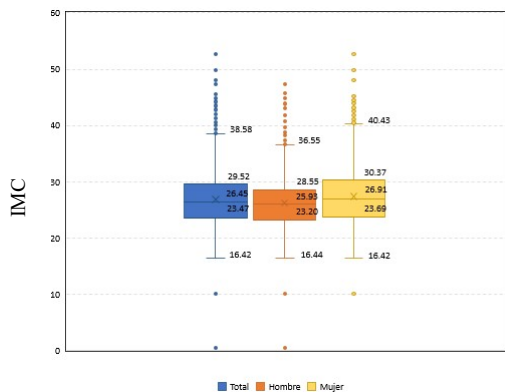


Fig. 1. Diagrama de caja del índice de masa corporal de adultos entre 19 y 59 años (elaboración propia con datos tomados de ENSANUT 2012).

En el diagrama de caja de la *Ilustración 1* se observa que, tanto en los hombres como en las mujeres, existen datos fuera de los bigotes, (en su mayoría en la parte superior) lo cual muestra que hay un porcentaje significativo de personas con sobrepeso y que en las mujeres la mediana del IMC aumenta con respecto a los hombres. La *Fig. 2* muestra más claramente como el IMC en las mujeres es superior al de los hombres en las edades entre 19 y 59 años. Si consideramos los grupos etarios estudiados, en las edades comprendidas entre (19-35) años, las mujeres presentan un 13% más casos de obesidad que los hombres. Mientras que en el grupo de (36-59) años, la diferencia cambia a 25% más de casos de obesidad en mujeres que en hombres.

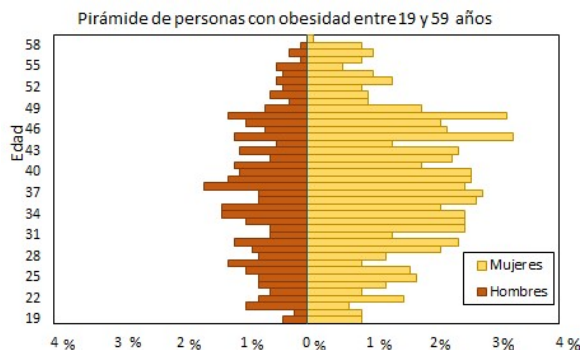


Fig. 2. Pirámide de adultos entre 19 y 59 años

Para establecer la asociación entre la obesidad y las variables explicativas se utilizó el estadístico chi cuadrado, con una significancia del 0.05. En cada una de las variables se estableció como grupo de referencia aquel que tuviera una menor probabilidad de padecer obesidad. A continuación, se presentan en la *Tabla I* las frecuencias de obesidad donde se muestran los grupos y probabilidades en los que la obesidad tiene mayor posibilidad de presentarse. Como se puede observar, se utiliza la prueba  $\chi^2$  que es una prueba de independencia a través del estadístico. Dado que el valor p es pequeño, indica asociación entre variables

TABLA I  
FRECUENCIA DE OBESIDAD

Variable	Presencia de obesidad						Valor p Chi Cuadrado	
	Si	%	No	%	Total	%		
Sexo	Hombre	320	16%	1641	84%	1961	100%	Pearson's Chisquared test Xsquared = 73.004, df = 1, p- value < 2.2e-16
	Mujer	721	27%	1953	73%	2674	100%	
	Total	1041	22%	3594	78%	4635	100%	
Grupo Edad	Adulto joven (19 - 35 años)	429	17%	2108	83%	2537	100%	Pearson's Chisquared test Xsquared = 98.424, df = 1, p- value < 2.2e-16
	Adulto (36 - 59)	612	29%	1486	71%	2098	100%	
	Total	1041	22%	3594	78%	4635	100%	
Provincia	Guayas	589	25%	1726	75%	2315	100%	Pearson's Chisquared test Xsquared = 23.294, df = 1, p- value = 1.39e-06
	Pichincha	452	19%	1868	81%	2320	100%	
	Total	1041	22%	3594	78%	4635	100%	
Área	Urbana	914	24%	2927	76%	3841	100%	Pearson's Chisquared test Xsquared = 22.546, df = 1, p- value = 2.051e-06
	Rural	127	16%	667	84%	794	100%	
	Total	1041	22%	3594	78%	4635	100%	

A partir de esto se realizó el modelo de regresión logística binaria, el cual nos permitirá evaluar la variable dicotómica (presencia o ausencia de la obesidad) y el impacto de cada una de las variables independientes (objeto de estudio). También se realizará el modelo de regresión Poisson, el cual nos permite determinar el número de personas con obesidad en cada provincia.

Para la construcción de los modelos de regresión se utilizó el método "introducción" o "enter", en el cual el proceso de selección de las variables es manual. Se parte de un modelo inicial, en el que se obliga a que entren todas las variables seleccionadas, se evalúa la variable que menos participa en él y se elimina, volviendo a construir un nuevo modelo de regresión aplicando la misma técnica, pero excluyendo la variable seleccionada y aplicando el mismo proceso de selección. Este proceso se repite reiteradamente hasta que se considere que el modelo obtenido es el que mejor se ajusta a las condiciones impuestas y que no se puede eliminar ninguna variable más de las que lo componen.

Una vez analizados los datos se procedieron a realizar los modelos a través del software estadístico R [12], suministrando como resultado los siguientes modelos ajustados:

### Modelo de Regresión Logística:

Aplicando regresión logística se obtiene los coeficientes de las variables grupo de edad, sexo, lugar de residencia y área (urbana o rural). Como se observa en la Fig. 3.

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.52789    0.11684  -13.077 < 2e-16 ***
grupo_edadAdultoJoven -0.69248    0.07245  -9.558 < 2e-16 ***
sexoMujer     0.64465    0.07633   8.446 < 2e-16 ***
provinciaPichincha -0.37382    0.07263  -5.147 2.65e-07 ***
AreaUrbana    0.48726    0.10597   4.598 4.27e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 244.040  on 15  degrees of freedom
Residual deviance: 24.238  on 11  degrees of freedom
    
```

Fig. 3. Coeficientes del modelo de regresión logística.

### Modelo ajustado:

$$\ln\left(\frac{p}{1-p}\right) = -1.53 - 0.69AdultoJoven + 0.64Mujer - 0.37Pichincha + 0.48Urbana$$

Se observa que todas las variables incluidas en el modelo son significativas con un residual deviance de 24.23 y un pseudo  $R^2=0.90$ . De este modo realizando el contraste chi-cuadrado se obtiene que todas las variables tienen un valor p significativo, es decir que su inclusión en el modelo permite explicar la obesidad.

```

Analysis of Deviance Table

Model: binomial, link: logit
Response: frec
Terms added sequentially (first to last)

              Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                15    244.040
grupo_edad  1    98.887      14    145.153 < 2.2e-16 ***
sexo        1    73.271      13    71.882 < 2.2e-16 ***
provincia   1    24.932      12    46.950 5.939e-07 ***
Area        1    22.712      11    24.238 1.882e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
    
```

Fig. 4. Tabla de análisis de deviance

### Bondad de ajuste

TABLA II  
PRUEBAS DE BONDAD DE AJUSTE

Prueba	Valor p
Hosmer-Lemeshow	1

Dado que la prueba de Hosmer-Lemeshow tiene un valor p cercano a 1, el ajuste presentado es bueno. Además, cuando el modelo es adecuado, los residuos tipo deviance deben ajustarse a una variable normal.

Al analizar los residuos del modelo a través de la prueba de Kolmogorov, se observa que no se rechaza la hipótesis nula de normalidad, indicando que los residuos si siguen una distribución normal.

```

One-sample kolmogorov-Smirnov test

data:  r1
D = 0.14387, p-value = 0.8496
alternative hypothesis: two-sided
    
```

Fig. 5. Prueba Kolmogorov

Es importante ver como clasifica el modelo, para ello se debe determinar el punto de corte para clasificar a alguien como obeso o no. Después de ajustar el modelo se determinó el punto de corte para la clasificación en 2.220446e-16, con cual se obtuvo la tabla de clasificación que se muestra en la figura 6.

```

> data$ajustado=floor(fitted(m2)+2.220446e-16)
> table(data$y,data$ajustado)

      0      1
0 3594      0
1  429     612
    
```

Fig. 6. Resultados de clasificación del modelo

De la tabla III se puede mencionar que el modelo de regresión binomial tiene un porcentaje de buena clasificación de 91%.

TABLA III  
PREDICCIÓN MODELO REGRESIÓN BINOMIAL

No. Obesos (Datos)	Pronóstico
89	69
11	9
94	95
16	19
128	137
25	19
194	207
32	29
45	46
3	7
54	64
8	7
113	122
15	15
197	169
17	17

**Modelo de Regresión Poisson:**

$$\log(\mu) = \beta_0 + \beta_1x_1 + \dots + \beta_px_p$$

```

Coefficients:
(Intercept)      Estimate Std. Error z value Pr(>|z|)
grupo_edadAdultoJoven -0.52195  0.06303  -8.282 < 2e-16 ***
sexoMujer        0.49164  0.06730   7.305 2.76e-13 ***
provinciaPichincha -0.27805  0.06266  -4.438 9.09e-06 ***
AreaUrbana       0.37649  0.09482   3.971 7.17e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 189.622  on 15  degrees of freedom
Residual deviance: 21.014  on 11  degrees of freedom
    
```

Fig. 7. Coeficientes del modelo de regresión Poisson

Aplicando regresión Poisson se obtienen los coeficientes de las variables grupo de edad, sexo, lugar de residencia y área (urbana o rural). Como se observa en la Fig. 7.

**Modelo ajustado:**

$$\log(\text{personas obesas}) = -1.75 - 0.52\text{AdultoJoven} + 0.49\text{Mujer} - 0.28\text{Pichincha} + 0.38\text{Urbana}$$

Se observa que todas las variables incluidas en el modelo son significativas con un residual deviance de 21.014 y un pseudo  $R^2=0.88$ . Esto se confirma, cuando se realiza el contraste chi-cuadrado de la Fig. 8, es decir que todas las variables explican a la obesidad.

```

Analysis of Deviance Table

Model: poisson, link: log
Response: num_personas_obesas
Terms added sequentially (first to last)

      Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                    15    189.622
grupo_edad  1    76.488      14    113.133 < 2.2e-16 ***
sexo        1    56.190      13    56.944 6.581e-14 ***
provincia   1    18.614      12    38.329 1.600e-05 ***
Area        1    17.316      11    21.014 3.165e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
    
```

Fig. 8. Tabla de análisis de Deviance

**Bondad de ajuste**

TABLA IV  
PRUEBAS DE BONDAD DE AJUSTE

Prueba	Valor p
Deviance	0.00
Chi cuadrado	0.03

El valor p de la prueba Deviance, indica que este modelo aún se aleja del modelo ideal. Por ello vamos a inspeccionar otros resultados que miden el ajuste del modelo.

Al analizar los residuos del modelo a través de la prueba de Kolmogorov, se observa que no se rechaza la hipótesis nula de normalidad, indicando que los residuos siguen una distribución normal, siendo un indicio de buen ajuste.

```

One-sample Kolmogorov-Smirnov test

data:  r3
D = 0.11466, p-value = 0.9685
alternative hypothesis: two-sided
    
```

Fig. 9. Prueba Kolmogorov

La Tabla V presenta los pronósticos del modelo:

TABLA V  
PREDICCIÓN MODELO REGRESIÓN POISSON

No. Obesos (Datos)	Pronóstico
89	70
11	9
94	92
16	18
128	134
25	19
194	214
32	29
45	49
3	8
54	63
8	7
113	121
15	15
197	168
17	17

Al comparar los dos modelos, aunque ambos producen buenas estimaciones, el modelo logístico es un poco mejor al modelo Poisson pues tiene un pseudo  $R^2$  más alto y una mejor predicción, por lo que se elegirá hacer la interpretación de los coeficientes del modelo de regresión binomial mostrados en la Tabla VI.

TABLA VI  
COEFICIENTES MODELO REGRESIÓN BINOMIAL

Variable	Valor coeficiente
Intercepto	$e^{-1.53} = 0.22$
Adulto Joven	$e^{-0.69} = 0.50$
Mujer	$e^{0.64} = 1.91$
Pichincha	$e^{-0.37} = 0.69$
Área Urbana	$e^{0.49} = 1.63$



El primer coeficiente indica que cuando el individuo es adulto, del sexo masculino y vive en el área rural de Guayas, entonces la posibilidad de que sea obeso disminuye en un 78%.

El segundo coeficiente indica que si un individuo es adulto joven y el resto de las variables permanecen constantes, entonces la posibilidad de ser obeso disminuye en un 50% que cuando el individuo es adulto.

El coeficiente de la variable sexo indica que la posibilidad de que un individuo sea obeso aumenta en un 91% cuando es mujer que cuando es hombre, manteniendo el resto de las variables constantes.

Con respecto al coeficiente de la variable provincia se puede decir que cuando un individuo pasa de vivir de la provincia de Guayas a la provincia de Pichincha manteniendo el resto constante, entonces la posibilidad de que sea obeso disminuye en un 31%.

Y finalmente el último coeficiente indica que, si un individuo pasa de vivir del área rural al área urbana con el resto de las variables constantes, entonces la posibilidad de que sea obeso aumenta en un 63%.

## V. CONCLUSIONES

En conclusión, realizando un análisis mediante una gráfica de pirámide poblacional (*Fig. 2*) se puede observar claramente como el IMC en la mujer es superior al de los hombres en las edades entre 19 y 59 años.

Mediante un análisis de descriptivo se verificó que un 27% de mujeres tiene prevalencia de obesidad en comparación a un 16% en hombres. También se observa que la provincia de Guayas (25%) tiene más personas con obesidad que la provincia de Pichincha (19%). En las personas adultas en edades entre 36 y 59 la obesidad es mayor de igual forma se manifiesta en las áreas urbanas y rurales.

Interpretando los coeficientes; todas las variables son significativas para los dos modelos que se estudiaron (Poisson y Binomial). En ambos modelos el valor de los coeficientes fueron los mismos, de acuerdo con los análisis realizados se muestra que la estructura de edad es un factor que afecta a la obesidad, mientras más aumente la edad, mayor será la prevalencia de obesidad en las personas.

Para el caso de la provincia también se observa un incremento de prevalecer este estado patológico, de igual forma sucede con el sexo de las personas, las mujeres predominan con este tipo de problema de salud. Los dos modelos resultaron con un buen ajuste para explicar la tasa de obesidad.

## REFERENCIAS

- [1] Smith, K. and Smith, M., «Obesity Statistics», *Primary Care: Clinics in Office Practice*, pp. 121-135, 2016.
- [2] Organización Mundial de la Salud, «Obesidad y Sobrepeso», 16 02 2018. [En línea]. Available: <https://www.who.int/es/news-room/fact-sheets/detail/obesity-and-overweight>.
- [3] INEC, «Tomo I: Encuesta Nacional de Salud y Nutrición de a población ecuatorianos de cero a 59 años. ENSANUT-ECU 2012.» Quito-Ecuador, 2014.
- [4] Bastos, A. A; González Boto, R.; Molinero González, O.; Salguero del Valle, A. «Obesidad, nutrición y actividad física.» *Revista Internacional de Medicina y Ciencias de Actividad Física y el Deporte*, pp. v5 n18 140-153, 2005.
- [5] Acosta-Hernández, M., Gasca-Pérez El., Ramos-Morales F.ernando, García –Rodríguez, R., Solís-Páez F., Evaristo-Portilla G., «Factores, causas y perspectivas de la obesidad infantil en México», *Medicas UIS*, Vol 26 Núm 1, pp. 59 - 68, 2013.
- [6] Dobson, A. «Introduction to generalized lineal models», Chapman & Hall /CRC, 2002
- [7] Menard, S., «Logistic Regression. From introductory to advanced concepts and applications», SAGE Publication, 2010.
- [8] Agresti Alan, «Categorical data analysis». Jhon Wiley & Sons, Inc. Publication, 2002
- [9] Kleinbaum, D., «Logistic Regression. A self-learning text», Atlanta Springer Science, 1994.
- [10] Cayuela, L., «Modelos lineales generalizados (GLM)», Granada, 2010.
- [11] Chau A., Lo E., Wong M., Chu CH., «Interpreting Poisson Regression Models in Dental Caries Studies», *Caries Research*, pp. 52(4), 339–345, 2018.
- [12] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria 2008. <https://www.R-project.org/>